Psychometrika

VOLUME XI — 1946
JANUARY-DECEMBER

Editorial Council

Chairman:—L. L. THURSTONE

Editors:—A. K. KURTZ
M. W. RICHARDSON

Managing Editor:—
HAROLD GULLIKSEN

Assistant Managing Editor:—
DOROTHY C. ADKINS

Editorial Board

H. S. CONRAD
ELMER A. CULLER
E. E. CURETON
JACK W. DUNLAP
MAX D. ENGELHART
HENRY E. GARRETT
J. P. GUILFORD
HAROLD GULLIKSEN
CHARLES M. HARSH
K. J. HOLZINGER
PAUL HORST
ALSTON S. HOUSEHOLDER
CLARK L. HULL
TRUMAN L. KELLEY
A. K. KURTZ
IRVING LORGE
QUINN MCNEMAR
CHARLES I. MOSIER
NICOLAS RASHEVSKY
M. W. RICHARDSON
P. J. RULON
CHARLES SPEARMAN
WM. STEPHENSON
S. A. STOUFFER
GODFREY THOMSON
L. L. THURSTONE
LEDYARD TUCKER
S. S. WILKS
HERBERT WOODROW

PUBLISHED QUARTERLY
By THE PSYCHOMETRIC SOCIETY
AT 23 WEST COLORADO AVENUE
COLORADO SPRINGS, COLORADO

Psychometrika

CONTENTS

MAXIMUM VALIDITY OF A TEST WITH EQUIVALENT ITEMS - - - - - - - - - - - - - - - 1
LEDYARD R TUCKER

FACTOR ANALYSIS AND BODY TYPES - - - - - - 15
L. L. THURSTONE

SERIAL CORRELATION - - - - - - - - - - - - 23
NATHAN JASPEN

A FACTORIAL ANALYSIS OF OLFACTION - - - - - 31
E. H. HSÜ

GRAPHICAL AND TABULAR AIDS FOR DETERMINING SAMPLE SIZE WHEN PLANNING EXPERIMENTS WHICH INVOLVE COMPARISONS OF PERCENTAGES - - - - - - - - - - - - - - - - 43
FRANCES SWINEFORD

A SIMPLE GRAPHICAL METHOD FOR ORTHOGONAL ROTATION OF AXES - - - - - - - - - - - - - 51
WAYNE S. ZIMMERMAN

A NEW DESCRIPTIVE STATISTIC: THE PARABOLIC CORRELATION COEFFICIENT - - - - - - - - - 57
CHARLES C. PETERS

VOLUME ELEVEN MARCH 1946 NUMBER ONE

A ten year topical index for *Psychometrika* (*1936-45 inclusive*) is being prepared. This will be issued in the near future as a separate folder which can be bound with Vol. 10.

PSYCHOMETRIKA—VOL. 11, NO. 1
MARCH, 1946

MAXIMUM VALIDITY OF A TEST WITH EQUIVALENT ITEMS

LEDYARD R TUCKER
COLLEGE ENTRANCE EXAMINATION BOARD

It is assumed that a scale of true scores on a function exists and that the probability of answering an item correctly is a curve of the type of the integral of the normal curve. The product moment correlation between the test score and true score is derived for a normal distribution of subjects and a test composed of equivalent items. Numerical examples demonstrate that the maximum correlation between test scores and true scores occurs for a one hundred item test when the point correlation between items is less than three tenths.

One of the most common assumptions in the theory of mental tests is the existence of a continuum of true scores which, with errors of measurement, produce the observed scores on a test. The errors of measurement are assumed to be unrelated not only to other errors of measurement, but also to the true scores; that is, for a particular true score, the errors have the same mean and standard deviation as for any other true score. Many of the formulas and concepts in the theory of mental tests are based on these two assumptions. Mental testers have used test reliabilities and corrections for attenuation in their thinking about test results for so long that these two concepts have become almost second nature to them. A commonly stated proposition is that any increase in the reliability of a test will increase the test's coefficients of validity. Since validity coefficients are difficult to obtain and work with, an intermediate goal of high test reliability has been established. New and improved types of items have been devised, more advantageous distributions of item difficulties have been investigated, and tests have been increased in length in order to increase the reliabilities of the tests. Although this work has resulted in better tests as indicated by increases in validity coefficients, critical consideration of several theoretical possibilities reveals inconsistencies between higher reliability and better measurement along the assumed continuum, and thus possible lower coefficients of validity.

Consider the case when all the items in a test are equivalent; that is, when the items all measure the same trait, have equal reliabilities, and are equally difficult. In this case the items are equally intercorrelated with coefficients equal to the item reliabilities. It is common experience that such a test with items of usual reliability administered to a random population produces a distribution of scores

which approaches a normal curve. If the reliability of the items were increased to unity, all correlations between the items would also become unity and a person passing one item would pass all items and another failing one item would fail all items. Thus the only possible scores are a perfect score or one of zero. Now the question can be raised whether the "measurement" obtained with the first test was due to the unreliability of the items or the results were solely due to a single imperfect discrimination. Is a dichotomy of scores the best that can be expected from a test with items of equal difficulty?

A resolution of these questions can be obtained when it is assumed that individuals with low true scores on the continuum of ability are less likely to answer each item correctly and individuals with high true scores are more likely to answer each item correctly. *Figure 1* shows how the probability of success on an item might be related to the true scores on the ability. The curve drawn on this figure represents an item that produces only a moderately good discrimination since the curve extends over a considerable range of ability scores. An item that gives a perfect discrimination would be represented by a vertical line located at such a score on the ability that all individuals with lower scores fail the item and all individuals with higher scores pass it. A completely invalid item would be represented by a horizontal line, for in this case the probability of passing an item is independent of an individual's score on the ability. Thus, if a single item is to be used, one represented by a vertical line would be preferred, since it would give perfect discrimination.

Consider the case, however, where a large number of equivalent items are to be used. In this case all of the items would be represented by the same curve. If the probability of correct response curve for these items were a vertical line, the total score over all items would accomplish nothing more than the perfect discrimination of any single item. Let the curve take a form such as illustrated in *Figure 1*, then a measure of the probability of success on an item could be translated into differential scores on the ability. If the total score on a test made up of a large number of such items is divided by the number of items, a measure of this probability of correct response for an item is obtained. The total score then produces a valid rank order of the individuals and a test with imperfect items gives better measurement of the scale of ability than a test with perfect items.

The foregoing proposition can be stated more precisely in mathematical terms. Item curves like the one in *Figure 1* can be thought of, in the simplest case, as depending on two parameters, one for the general level of difficulty and another for the discriminative power of the item. A convenient index of an item's difficulty is the score on

the scale of ability where the probability of correct responses is one half and will be designated by s_j. The parameter for an item's discriminative power can be described in terms of a coefficient of the amount of spread of the item curve. This coefficient, which will be designated σ_j, has an inverse relation to the discriminative power of the item in that a large σ_j indicates a flatter curve than a small σ_j and thus a lower discriminative power. An item which gave a perfect discrimination would have a σ_j equal to zero. Equivalent items will have equal s_j's and σ_j's. The correlation, r_{ts}, of a test made of n of these items with the scale of ability is, for a given population, a function of s_j, σ_j, and n as stated below:

$$r_{ts} = \phi(s_j, \sigma_j, n). \tag{1}$$

Since it is desired to investigate the relation of r_{ts} and σ_j, it can be assumed that s_j and n are constants. This relation will be developed in the following section for the special case where the population has a normal distribution on the scale of ability; but, since most populations tested approximate such a distribution, the results should hold approximately for most situations.

Following is a summary of the notation that will be used.

j = a subscript indicating an item.

k = a subscript for an alternative item.

t = a subscript indicating the test.

s = both the subscript for the scale of ability and a score on the scale of ability.

p_{js} = the probability of a correct response to item j by an individual with ability score s.

s_j = the s where pjs is one half.

σ_j = a constant of the pjs, s curve which determines the spread of this curve.

N = number of individuals within the population tested.

f_s = the frequency of individuals with ability scores of s within the population tested.

f_{js} = the frequency of individuals with ability score s who pass item j.

p_j = the proportion of the individuals in the population who give the correct response to item j.

f_{jks} = the frequency of individuals with ability score s who pass both items j and k.

p_{jk} = the proportion of the individuals in the population who pass both items j and k.

r_{js} = the product moment correlation coefficient between item j and the scale of ability s.

r_{jk} = the product moment correlation coefficient between item j and item k.

r_{ts} = the product moment correlation coefficient between test t and the scale of ability s.

n = the number of items in the test.

Several arbitrary variables will be introduced during the mathematical development and their designation will be explained in the context.

In order to give substance to the development it is necessary to assume an equation for the curve of *Figure 1*, and, since its general

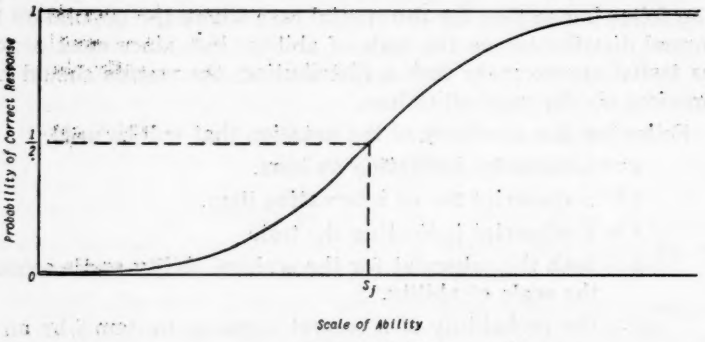


FIGURE 1

shape is the same as the integral of the normal curve, the assumption is made that it can be represented by the following equation:

$$p_{js} = \int_{u_j=-\infty}^{u_j=(s-s_j)/\sigma_j} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u_j^2}\, du_j. \tag{2}$$

The arbitrary variable u_j is used in order not to confuse this integration with other operations on the s scale.

A second necessity is to have a population to work with; therefore, it is assumed that the distribution of scores on the s scale is normal with a unit standard deviation and a mean of zero.

$$f_s = \frac{N}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2}. \tag{3}$$

The first task of the development is to derive the equation for the frequency of individuals with score s who pass item j. Using the probability of correct response p_{js} as the proportion of the individuals with score s who pass the item, the following equation can be written:

$$f_{js} = f_s \, p_{js} . \tag{4}$$

Substituting from equations (2) and (3), equation (4) becomes:

$$f_{js} = \frac{N}{2\pi} \int_{u_j=-\infty}^{u_j=(s-s_j)/\sigma_j} e^{-\frac{1}{2}(s^2+u^2)} \, du_j . \tag{5}$$

This equation for f_{js} is difficult to work with since s occurs as one of the limits of integration. This difficulty is eliminated by the following transformation of variables, s being held constant:

Then:

$$v_j \equiv (s - \sigma_j u_j) \frac{1}{\sqrt{1+\sigma_j^2}} . \tag{6}$$

$$u_j = \frac{1}{\sigma_j} [s - v_j \sqrt{1+\sigma_j^2}] , \tag{7}$$

$$du_j = -\frac{1}{\sigma_j} \sqrt{1+\sigma_j^2} \, dv_j . \tag{8}$$

When this transformation is performed, equation (5) becomes:

$$f_{js} = -\frac{N\sqrt{1+\sigma_j^2}}{2\pi\sigma_j} \int_{v_j=\infty}^{v_j=s_j/\sqrt{1+\sigma_j^2}} e^{-\frac{1}{2}A} \, dv_j , \tag{9}$$

where

$$A = \frac{1+\sigma_j^2}{\sigma_j^2} \left(s^2 + v_j^2 - 2 \frac{s \, v_j}{\sqrt{1+\sigma_j^2}} \right) .$$

It is now advantageous to define:

$$r_{vs} \equiv \frac{1}{\sqrt{1+\sigma_j^2}} . \tag{10}$$

When equation (10) is substituted in (9) and the limits of integration are interchanged accompanied with a change in algebraic sign, the equation for f_{js} becomes:

$$f_{js} = \frac{N}{2\pi\sqrt{1-r_{vs}^2}} \int_{v_j = r_{vs} s_j}^{v_j = \infty} e^{-\frac{1}{2}B}\, dv_j , \tag{11}$$

where

$$B = \frac{1}{1-r_{vs}^2}\,(s^2 + v_j^2 - 2\, s\, v_j\, r_{vs}) .$$

This equation gives the interesting fact that the variables v_j and s have a normal surface of frequency with a correlation r_{vs}. The distribution on v_j is also a unit normal curve with the lower limit of integration being the value of v_j which is predicted from s_j by the regression of v_j on s.

The proportion of individuals in the population who pass item j can be obtained by integrating f_{js} for all values of s as follows:

$$p_j = \frac{1}{N} \int_{s=-\infty}^{s=\infty} f_{js}\, ds . \tag{12}$$

When the right side of equation (11) is substituted for f_{js} in equation (12),

$$p_j = \frac{1}{2\pi\sqrt{1-r_{vs}^2}} \int_{v_j = s_j r_{vs}}^{v_j = \infty} \int_{s=-\infty}^{s=\infty} e^{-\frac{1}{2}B}\, dv_j\, ds. \tag{13}$$

This equation integrates to:

$$p_j = \int_{v_j = s_j r_{vs}}^{v_j = \infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v_j^2}\, dv_j . \tag{14}$$

The proportion passing an item can then be found from equations (10) and (14) for any values of σ_j and s_j.

The next problem to be considered is the product moment correlation of an item with the scale of ability. Let the scores on an item be designated by x_j. These scores are either 0 or 1. Equation (15) gives the formula for the correlation of item j with ability s:

$$r_{js} = \frac{\Sigma\, x_j\, s}{N\sqrt{p_j\, q_j}} . \tag{15}$$

The mean and σ of the s's do not enter in this formula since they are zero and unity respectively. Let the s's for individuals with unit scores on the item be designated s_1; then:

$$\Sigma\, x_j\, s = \Sigma\, s_1 . \tag{16}$$

The individuals with x_j's of zero do not contribute to this sum. When this summation is expressed in integral form it becomes:

$$\Sigma s_1 = \int_{s=-\infty}^{s=\infty} f_{js}\, s\, ds. \tag{17}$$

Substitution for f_{js} its value in equation (11) this becomes:

$$\Sigma s_1 = \frac{N}{2\pi\sqrt{1-r_{vs}^2}} \int_{v_j=s_j r_{vs}}^{v_j=\infty} \int_{s=-\infty}^{s=\infty} s\, e^{-\frac{1}{2}B}\, ds\, dv_j\,, \tag{18}$$

which integrates to:

$$\Sigma s_1 = \frac{N r_{vs}}{\sqrt{2\pi}} e^{-\frac{1}{2}(s_j r_{vs})^2}. \tag{19}$$

Define:

$$y_{(s_j r_{vs})} \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(s_j r_{vs})^2}. \tag{20}$$

Then substituting in equation (15):

$$r_{js} = r_{vs} \frac{y_{(s_j r_{vs})}}{\sqrt{p_j q_j}}. \tag{21}$$

This relation is not unexpected since r_{vs} is the coefficient of the normal surface between v_j and s and is therefore the biserial coefficient of correlation. r_{js} is the point biserial coefficient. Equation (21) is the formula relating these two coefficients.

As a preliminary to deriving the formula for the intercorrelation between two items, j and k, it is necessary to find the equation for the proportion of the individuals who pass both items. Since the two items are assumed to be related only by the single ability s, the probability, p_{jks}, of individuals with the same ability passing both items can be written by the law of joint probability:

$$p_{jks} = p_{js}\, p_{ks}\,, \tag{22}$$

and:

$$f_{jks} = f_s\, p_{js}\, p_{ks}\,. \tag{23}$$

Substituting equations (2) and (3) into equation (23):

$$f_{jks} = \frac{N}{(2\pi)^{3/2}} \int_{u_j=-\infty}^{u_j=(s-s_j)/\sigma_j} \int_{u_k=-\infty}^{u_k=(s-s_k)/\sigma_k} e^{-\frac{1}{2}(u_j^2+u_k^2+s^2)}\, du_j\, du_k\,. \tag{24}$$

The transformation of variables of equation (6), (7), and (8) is used to remove s from the limits of integration. Equation (24) then becomes:

$$f_{jks} = \frac{N\sqrt{1+\sigma_j^2}\,\sqrt{1+\sigma_k^2}}{(2\pi)^{3/2}\,\sigma_j\,\sigma_k} \int_{v_j=s_j r_{vs}}^{v_j=\infty} \int_{v_k=s_k r_{vs}}^{v_k=\infty} e^{-iC}\,dv_j\,dv_k\,, \qquad (25)$$

where

$$C = \frac{s^2}{\sigma_j^2} + \frac{(1+\sigma_j^2)}{\sigma_j^2} v_j^2 - 2\,s\,v_j \frac{\sqrt{1+\sigma_j^2}}{\sigma_j^2} + \frac{s^2}{\sigma_k^2} + \frac{(1+\sigma_k^2)}{\sigma_k^2} v_k^2 - 2\,s\,v_k \frac{\sqrt{1+\sigma_k^2}}{\sigma_k^2}.$$

Since the items are equivalent

$$\sigma_j = \sigma_k\,, \qquad (26)$$

$$s_j = s_k\,; \qquad (27)$$

r_{vs} of equation (10) will be used for both items: Equation (25) then becomes:

$$f_{jks} = \frac{N}{(2\pi)^{3/2}\,(1-r_{vs}^2)} \int_{v_j=r_{vs}s_j}^{v_j=\infty} \int_{v_k=r_{vs}s_j}^{v_k=\infty} e^{-iD}\,dv_j\,dv_k\,, \qquad (28)$$

where

$$D = \frac{1+r_{vs}^2}{1-r_{vs}^2} s^2 + \frac{1}{1-r_{vs}^2}(v_j^2+v_k^2) - 2\frac{r_{vs}}{1-r_{vs}^2} s(v_j+v_k)\,.$$

In order to find the proportion, p_{jk}, who pass both items, f_{jks} is integrated over all values of s:

$$p_{jk} = \int_{s=-\infty}^{s=\infty} \frac{1}{N} f_{jks}\,ds\,. \qquad (29)$$

Substituting the value of f_{jks} in equation (28) into equation (29) gives:

$$p_{jk} = \frac{1}{(2\pi)^{3/2}\,(1-r_{vs}^2)} \int_{s=-\infty}^{s=\infty} \int_{v_j=r_{vs}s_j}^{v_j=\infty} \int_{v_k=r_{vs}s_j}^{v_k=\infty} e^{-iD}\,dv_j\,dv_k\,ds\,. \qquad (30)$$

The terms involving s can be collected, the square completed, and the integration over s performed so as to obtain:

$$p_{jk} = \frac{1}{2\pi\sqrt{1-r_{v_jv_k}^2}} \int_{v_j=r_{vs}s_j}^{v_j=\infty} \int_{v_k=r_{vs}s_j}^{v_k=\infty} e^{-iE}\,dv_j\,dv_k\,, \qquad (31)$$

where

$$E = \frac{1}{(1 - r^2_{v_j v_k})} (v_j^2 + v_k^2 - 2 v_j v_k r_{v_j v_k}) ;$$

and

$$r_{v_j v_k} = r_{vs}^2 , \tag{32}$$

or

$$r_{v_j v_k} = \frac{1}{1 + \sigma_j^2} . \tag{33}$$

As in the case of the relation v_j and s the bivariate distribution between v_j and v_k is a normal surface, but in this case cut by two planes, one parallel to each of the axes. The proportions in the four quadrants resulting from this double division by the two planes are the entries in the four-fold correlation table which has $r_{v_j v_k}$ as its tetrachoric coefficient. A further point of interest about $r_{v_j v_k}$ is that when the effect of the ability s is partialled out, the coefficient of partial correlation $r_{v_j v_k . s}$ is zero:

$$r_{v_j v_k . s} = \frac{r_{v_j v_k} - r_{vs}^2}{\sqrt{1 - r_{vs}^2} \sqrt{1 - r_{vs}^2}} = 0 . \tag{34}$$

When the product moment coefficient of correlation between the two items is found, as indicated in the following section, the partial correlation found from it is not zero.

Equation (35) is the formula for the product moment correlation between item j and k:

$$r_{jk} = \frac{p_{jk} - p_j p_k}{\sqrt{p_j q_j} \sqrt{p_k q_k}} . \tag{35}$$

Since the items are equivalent,

$$p_j = p_k \tag{36}$$

and

$$r_{jk} = \frac{p_{jk} - p_j^2}{p_j q_j} . \tag{37}$$

The next problem is to obtain an equation relating the scores on a test composed of n items and the ability s. The formula for the correlation between the test scores, x_t, and scores on ability s is:

$$r_{ts} = \frac{\Sigma x_t s}{N \sigma_t} . \tag{38}$$

The variance of the scores on the test can be found by equation (39):

$$\sigma_t^2 = [n + n(n-1)r_{jk}]\, p_j\, q_j \tag{39}$$

since the variance of the individual items is $p_j\, q_j$ and they are equally intercorrelated. The numerator of equation (38) can be written as:

$$\Sigma\, x_t\, s = \Sigma_j (\Sigma\, x_j) s \tag{40}$$

which becomes:

$$\Sigma\, x_t\, s = \Sigma_j \Sigma\, x_j\, s, \tag{41}$$

or by equation (15) and the fact that the items are equivalent:

$$\Sigma\, x_t\, s = N\, n\, r_{js}\, \sqrt{p_j\, q_j}\,. \tag{42}$$

When equation (39) and (42) are substituted in (38)

$$r_{ts} = \frac{n\, r_{js}}{\sqrt{n + n(n-1)r_{jk}}}. \tag{43}$$

It is now possible to start with given values of s_j, σ_j and n compute the correlation of the test with the ability. The important computational steps are to find the values of:

1) r_{vs} from equation (10),

2) p_j from equation (14),

3) r_{js} from equation (21),

4) $r_{v_j v_k}$ from equation (32),

5) p_{jk} from equation (31),

6) r_{jk} from equation (37),

7) r_{ts} from equation (43).

Figure 2 shows the relation of the correlation, r_{ts}, of the scores on the test with those on the ability and the major dependent variable σ_j when s_j is zero and the test is composed of 1 item, 10 items, or 100 items. Only in the case when the test has a single item does r_{ts} have its highest value when σ_j is zero, in each of the other cases the maximum r_{ts} occurs for a finite value of σ_j greater than zero. When a test is composed of more than one item its validity increases to a maximum value as σ_j increases from zero and then falls off again for less diagnostic items. *Figure 3* shows that this same relation holds for tests whose items are not pitched at the mean of the population.

For illustrative purposes a value of s_j equal to unity was chosen. Similar curves could be drawn for items of other difficulties.

The relation indicated above is given in more conventional form in *Figures 4* and *5* where the correlation of the test with the ability is plotted against the intercorrelations of the items. If the scale of ability is taken as the criterion, r_{ts} is the coefficient of validity. The item intercorrelation, r_{jk}, can be interpreted as the reliability coefficient of the items. *Figure 4* shows that a one hundred item test, perfectly pitched in difficulty for the population tested, has a validity of slightly less than .8 for items with perfect reliability, but attains a maximum validity over .97 when the items have a reliability slightly over .2.

A solution for the σ_j which would give a maximum r_{ts} for fixed values of s_j and n would be of great interest but cannot be obtained in the general case since the equations (14) and (31) cannot be integrated except in special cases. The results illustrated in *Figures 2-5*, however, substantiates the conclusion that, for tests composed of items of equal difficulty, best measurement is obtained when these items have less than perfect discriminative power and reliability.

A result which seemed amazing was the low values of the item reliabilities which yielded the best measurement. The item reliabilities for maximum test vs. ability correlations were within the range of practical experience. This fact would not, however, necessarily indicate that item reliabilities should be held down to the values indicated in the Figures. Since the curves rise sharply for low item reliabilities and fall slowly for items with higher reliabilities, and since item reliabilities are unstable, it is safer for the reliabilities to be too high.

Although the preceding development is of interest in pointing out an inconsistency between high test reliability and good measurement, most tests are constructed with a distribution of item difficulties; therefore, the assumption of equivalent items does not hold and the results are not directly applicable. It is hoped that an extension of the present system will indicate the validity properties of a test when the items differ in difficulty and that a system can be developed to determine the best distribution of item difficulties.

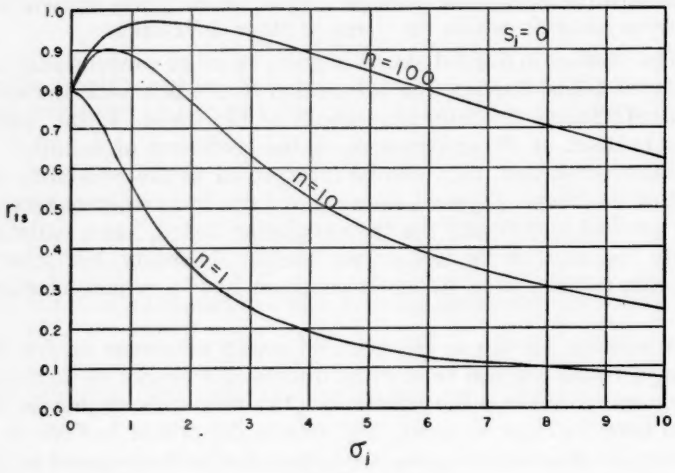


FIGURE 2

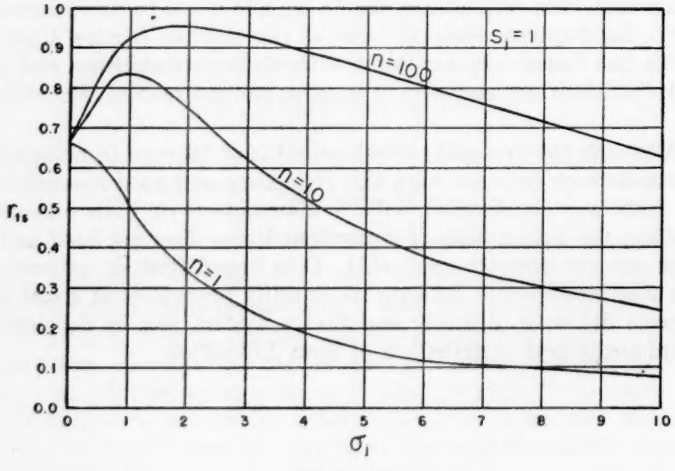


FIGURE 3

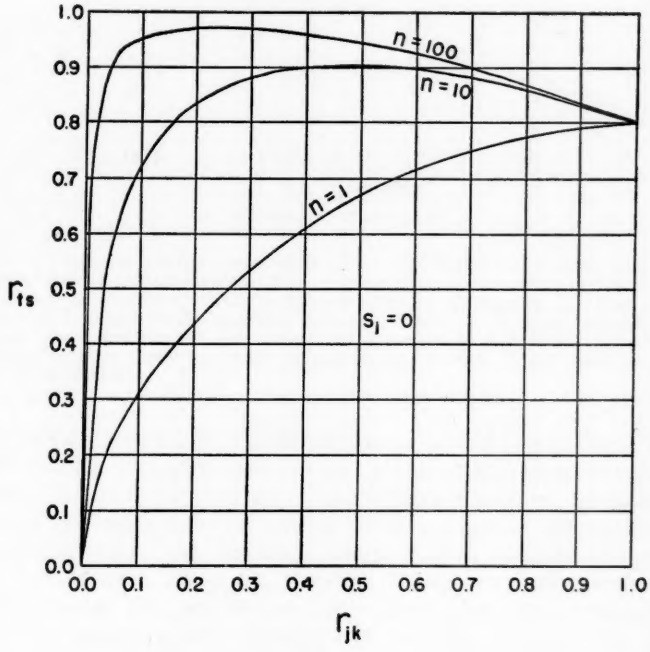


FIGURE 4

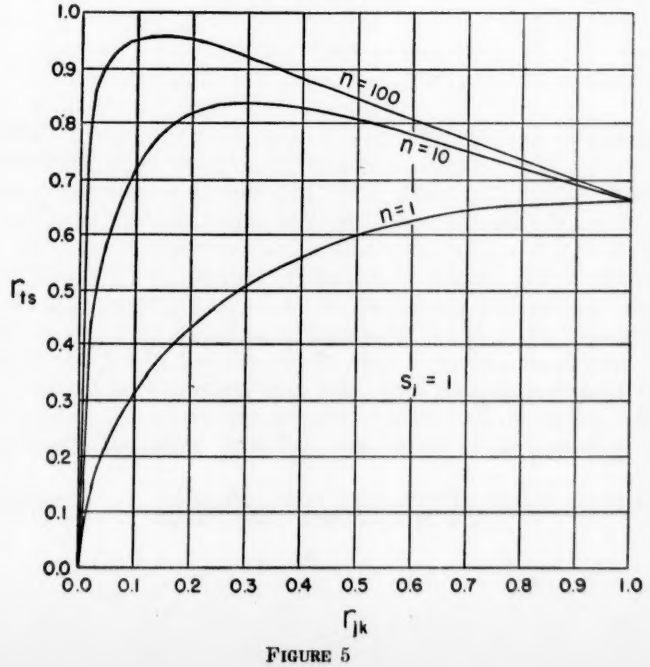
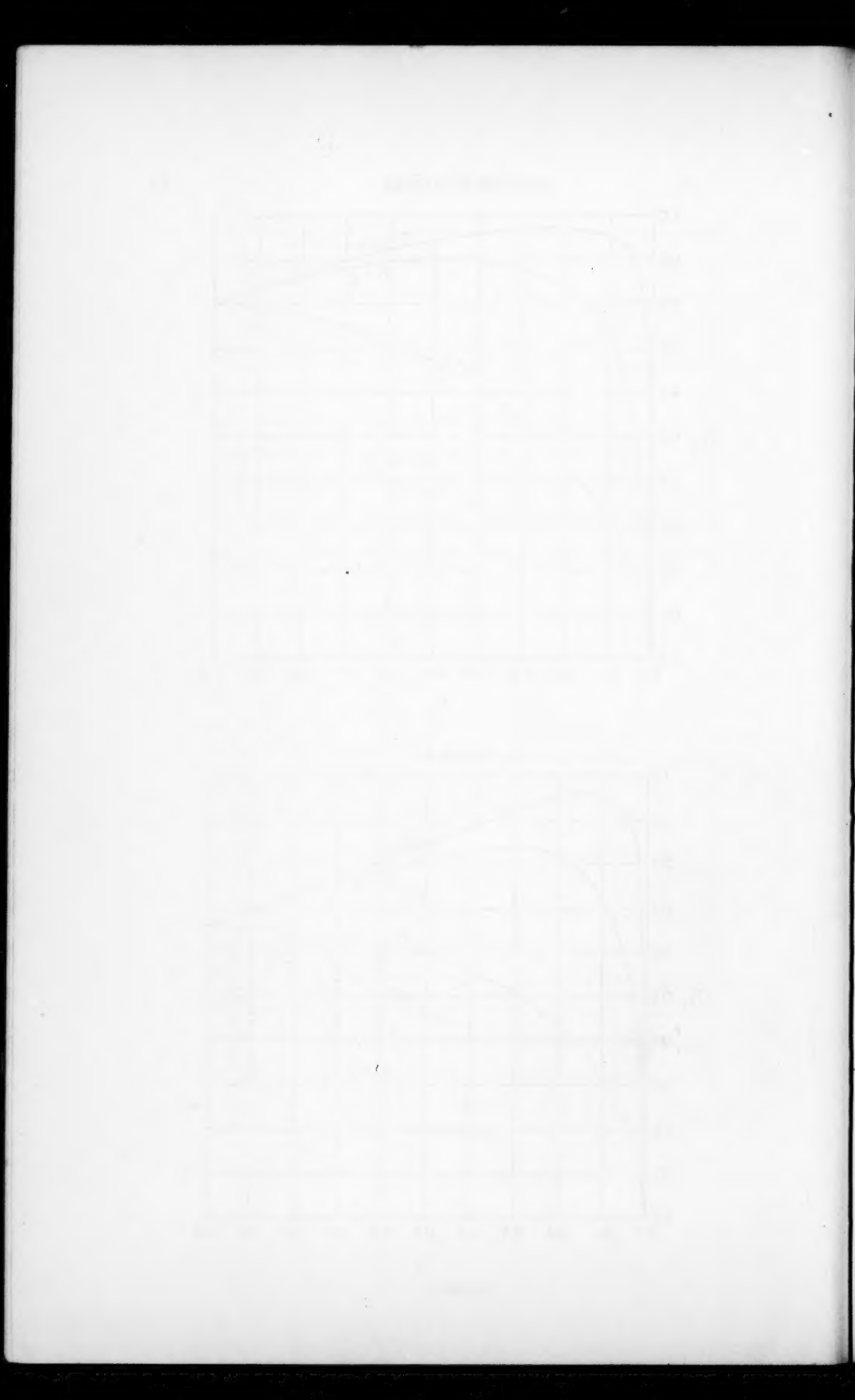


FIGURE 5



PSYCHOMETRIKA—VOL. 11, NO. 1
MARCH, 1946

FACTOR ANALYSIS AND BODY TYPES

L. L. THURSTONE
THE UNIVERSITY OF CHICAGO

A factorial analysis was made of a small battery of twelve anthropometric measurements. The correlations can be accounted for by four factors in a simple structure. This small battery has been used by the author for teaching purposes. Several of these factors seem to be meaningful, but their acceptance must depend on more comprehensive studies of body measurements with a larger number of measurements.

In a recent paper* W. H. Hammond of University College, London, described a factorial analysis of twelve body measurements on adult men. After computing all of the intercorrelations, he factored the correlation matrix by a method of Cyril Burt. Hammond attempted to interpret the resulting factor matrix directly and without rotation of axes. It is the purpose of this paper to present an alternative factor analysis of the same correlation matrix which reveals an interesting simple structure with rotation of axes. The writer has used Hammond's data for teaching purposes because it is unusual for a small battery of twelve variables to give so clear a simple structure with rather easily interpreted primary factors and with rank as low as 4. Because of the relative simplicity of this alternative factorial interpretation it may help to clarify the nature and possibilities of the simple structure concept.

In dealing with the body type problem by factorial methods we did not consider types to be in the nature of discrete classes by which each individual may be categorized. It is doubtful whether students of the body type problem really believe in discrete categories without overlap. It seems more profitable to deal with the body type problem with a relatively small number of factors or coordinates in terms of which each individual physique can be described. If a person belongs to a type his physique would be conspicuous in one of the descriptive coordinates and not in the others. In such a coordinate system each individual might be described as a point in a descriptive space of as many dimensions as there are parameters, factors, or types, whichever they might be called. It is then a question of fact whether indi-

* Hammond, W. H. An application of Burt's multiple general factor analysis to the delineation of physical types. *Man*, 1942, 42, 4-11.

viduals exist who would be described by points on the coordinate axes or in the coordinate planes. For some factors this might be physically possible whereas for others it might not be possible.

In *Table 1* we have reproduced Hammond's correlation matrix for the twelve anthropometric measurements. It was factored by one of the centroid methods which produced the orthogonal factor matrix F_0 in *Table 2*. The absolute mean value of the fourth-factor residuals was .0311. Because the residuals were vanishingly small it was inferred that four factors were adequate to account for the variance in the given correlation matrix.

The numerical values in the factor matrix F_0 of *Table 2* are determined by the method of factoring that happens to be used. The particular factoring method is here of secondary consequence provided that it produces a minimum distortion of the configuration in the common factors. Any legitimate factor method could be used because the primary factors which are to be interpreted are determined by the test configuration and not by the factoring method.

Pairs of columns of F_0 were plotted and four successive adjustments were made in the rotation of axes. The transformation matrix Λ is shown in *Table 3*. This transformation matrix carries the given factor matrix F_0 to the oblique factor matrix V of *Table 4*. We have then the matrix equation $F_0\Lambda = V$. The matrix $C = \Lambda'\Lambda$ is given in *Table 3*, and it shows the cosines of the angular separations between the reference factors. The configuration of the oblique factor matrix V is represented in the six diagrams of the accompanying figure. These six diagrams indicate no further adjustment in the four reference axes A, B, C, D. The analysis of the final set of diagrams is the most important and interesting part of a factorial study.

In the first diagram we have a plot of the projections of B against those of A. These are obtained directly from the columns A and B of *Table 4*. The two lines of this diagram represent the unit reference vectors A and B. All of these diagrams are drawn in orthogonal form for convenience rather than in oblique form because the structure is unaffected by this procedure except when the separating angles become quite small. The diagram BA shows three measures with significant saturation on the reference axis A. These are measures 8, 9, 10. In this study we have regarded saturations as significant if the factor loadings exceed .3 or .35, which mean that at least one-tenth of the variance of the measurement is attributable to the factor. Referring to the list of anthropometric measures in *Table 1* we find that these measures are head length, head breadth, and head height. The diagram BA shows that all of the other measures have vanishingly small projections on the reference axis A. We find that there are only

three measures of head size and that all of them have significant saturation on A while all other measures have vanishingly small projections on this axis. This characteristic of the configuration leads us to the hypothesis that in the development of body proportions there exists a factor or parameter that is concerned with head size.

In the same diagram BA we find that there are four measures with significant saturation on the reference axis B and that all of the other measures have vanishingly small projections on this axis. These four measures are stature, sitting height, span, and hand length. That which is common to these four measures seems to be bone length, and one is justified to formulate the hypothesis that in the development of body proportions some factor or parameter is primarily concerned with bone length.

The diagram CA shows four measures with significant saturation on the reference axis C. These are measures 3, 4, 6, 7; namely, shoulder breadth, hip breadth, chest breadth, and chest depth. That which is common to these measures seems to be trunk size or girth. This characteristic of the configuration leads us to postulate a parameter or factor that is primarily concerned with transverse measures, which may be denoted girth. They are principally concerned with transverse measures of the trunk.

The diagram DA shows four measures which have significant saturation on the reference axis D whereas all of the other measures have small projections on that axis. These are measures 1, 5, 11, 12; namely, stature, span, hand length, and hand breadth. These four measures include all of the measures of the extremities in this small battery of twelve variables. One might therefore postulate that there is a factor or parameter primarily concerned with size of extremities. The diagram DB is of special interest in our attempt to differentiate between the two factors D and B. In that diagram we see that the three variables 1, 5, 11 have significant saturations in both B and D but the situation is quite different with the measures 2 and 12; namely, sitting height and hand breadth. As far as this small battery of twelve variables is concerned, the differentiation between D and B hinges on the separation between sitting height and hand breadth. The factor B has a significant effect on sitting height but not on hand breadth. On the other hand the factor D has a significant effect on hand breadth but not on sitting height.

The differentiation between these two factors, if they should be sustained in subsequent studies of larger batteries, would depend on the inclusion of additional measures which should represent a crucial separation between these two postulated factors. If several rival hypotheses should be formulated, then additional measures should be in-

cluded in a factorial study which would crucially differentiate between the several hypotheses in the hope that one of them might be sustained. That can be done only by analyzing a larger set of body measurements.

In *Table 5* we have the correlations between the four primary factors that correspond to the four reference axes. It will be seen that these four primary factors or parameters are essentially uncorrelated, except *B* and *C*, which show a correlation of .38. These two primary factors have here been given the tentative interpretation length and girth, respectively. One might infer, therefore, that length and girth have a significant positive correlation, which is what we should probably expect. On the other hand, head size and the size of extremities seem to be two relatively independent factors.

Table 6 shows the oblique factor pattern of this battery. In this table each significant factor loading of *Table 4* is represented by an "*x*" and the blank cells have vanishing saturations that can be ignored. This is an alternative representation of the simple structure which is useful for inspectional purposes. It shows at a glance, for example, that, as regards the common factors of this battery, stature can be represented as a linear combination of factors *B* and *D*.

Summarizing this simple analysis, we have the following findings which should be regarded as tentative and subject to more intensive analysis on larger batteries.

(1) The factors that determine body size and proportions seem to constitute a system that is simpler than the factors which determine the cognitive or intellective functions. We make this inference because similar factorial studies of the cognitive functions seldom reveal so clear a structure with only twelve tests and four factors.

(2) The present battery seems to indicate the existence of four primary factors or parameters which are concerned with the development of: (a) head size, (b) bone length, (c) transverse measures or girth, (d) a factor less clear but which may be concerned with size of extremities.

This paper will have served its main purpose if it should be provocative of more intensive studies of larger sets of anthropometric measurements with the aid of the simple structure principle. There is good indication that it should be possible to identify those parameters or factors which operate to produce the individual variations in body

size and proportions. It is to be hoped that the factors that may be so identified will be directly related to the work of anatomists and endocrinologists toward a scientific understanding of the mechanism of body growth.

TABLE 1
Correlation Matrix R

		1	2	3	4	5	6	7	8	9	10	11	12
1.	Stature	.81	.72	.24	.27	.81	.21	.06	.25	.04	.09	.69	.41
2.	Sitting height	.72	.77	.23	.23	.51	.29	.11	.16	.02	.16	.55	.15
3.	Shoulder breadth	.24	.23	.41	.34	.41	.48	.31	.19	.06	.06	.30	.28
4.	Hip breadth	.27	.23	.34	.53	.39	.53	.47	.26	.04	−.04	.35	.42
5.	Span	.81	.51	.41	.39	.81	.33	.12	.22	.15	.11	.74	.51
6.	Chest breadth	.21	.29	.48	.53	.33	.53	.36	.33	.13	−.02	.30	.37
7.	Chest depth	.06	.11	.31	.47	.12	.36	.41	.10	.04	−.03	.10	.20
8.	Head length	.25	.16	.19	.26	.22	.33	.10	.33	.32	.32	.22	.16
9.	Head breadth	.04	.02	.06	.04	.15	.13	.04	.32	.35	.35	.23	−.01
10.	Head height	.09	.16	.06	−.04	.11	−.02	−.03	.32	.35	.35	.11	.03
11.	Hand length	.69	.55	.30	.35	.74	.30	.10	.22	.23	.11	.71	.63
12.	Hand breadth	.41	.15	.28	.42	.51	.37	.20	.16	−.01	.03	.63	.63

TABLE 2
Orthogonal Factor Matrix F_0

	I	*II*	*III*	*IV*	h^2
1.	.72	—.55	—.15	—.11	.8555
2.	.60	—.34	—.25	—.48	.7685
3.	.51	.17	.22	—.06	.3410
4.	.59	.21	.36	—.11	.5339
5.	.79	—.35	.02	.03	.7479
6.	.60	.32	.24	—.16	.5456
7.	.34	.34	.29	—.26	.3829
8.	.45	.31	—.30	.04	.3902
9.	.28	.29	—.36	.32	.3945
10.	.23	.16	—.49	.19	.3547
11.	.78	—.41	.01	.23	.8295
12.	.58	—.15	.38	.31	.5994

TABLE 3

	Transformation Matrix Λ					$C = \Lambda'\Lambda$			
	A	*B*	*C*	*D*		*A*	*B*	*C*	*D*
I	.27	.40	.47	.46	*A*	1.00	—.14	.00	—.07
II	.54	—.64	.64	—.45	*B*	—.14	1.00	—.38	—.01
III	—.70	—.42	.58	.32	*C*	.00	—.38	1.00	—.01
IV	.39	—.50	—.17	.70	*D*	—.07	—.01	—.01	1.00

TABLE 4

Oblique Factor Matrix *V*

	A	*B*	*C*	*D*
1.	—.04	.76	—.08	.45
2.	—.03	.80	.00	.01
3.	.05	.03	.49	.19
4.	—.02	.01	.64	.22
5.	.02	.52	.15	.55
6.	.10	.01	.65	.10
7.	—.03	—.07	.59	—.09
8.	.51	.09	.23	.00
9.	.61	—.08	.05	.11
10.	.57	.10	—.11	.01
11.	.07	.46	.07	.71
12.	—.07	.01	.34	.67

TABLE 5

Correlations of Primary Factors

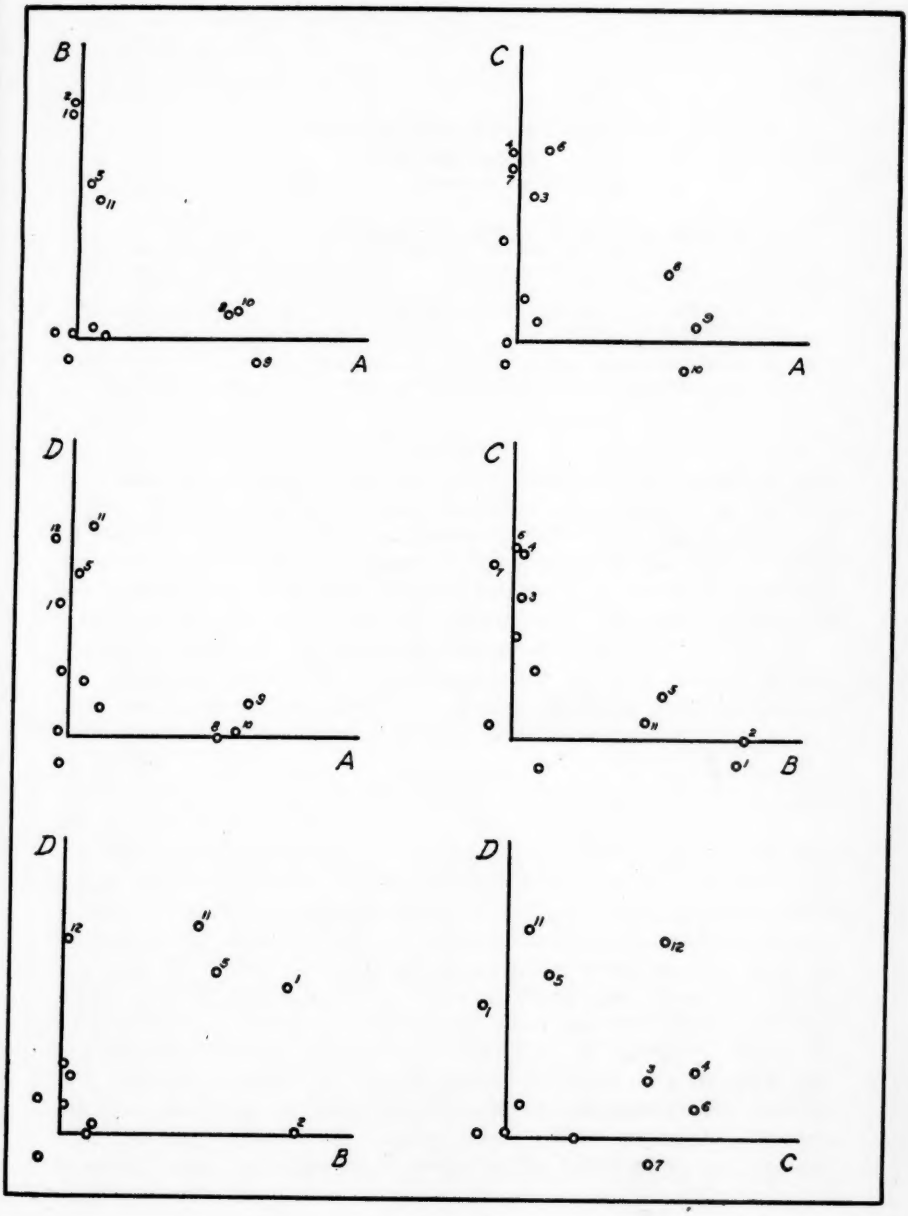
$R_{pq} = D(\Lambda'\Lambda)^{-1}D$

	T_A	T_B	T_C	T_D
T_A	1.00	.16	.06	.07
T_B	.16	1.00	.38	.03
T_C	.06	.38	1.00	.02
T_D	.07	.03	.02	1.00

TABLE 6

Factor Pattern

		A	*B*	*C*	*D*
1.	Stature		x		x
2.	Sitting height		x		
3.	Shoulder breadth			x	
4.	Hip breadth			x	
5.	Span		x		x
6.	Chest breadth			x	
7.	Chest depth			x	
8.	Head length	x			
9.	Head breadth	x			
10.	Head height	x			
11.	Hand length		x		x
12.	Hand breadth			(x)	x


B
A
C
A
D
A
C
B
D
B
D
C

PSYCHOMETRIKA—VOL. 11, NO. 1
MARCH, 1946

SERIAL CORRELATION

NATHAN JASPEN
UNITED STATES EMPLOYMENT SERVICE

Formulas are presented for triserial correlation, quadriserial correlation, etc., and for serial correlation in general. These formulas are based on well-known procedures outlined by Kelley, Peters and Van Voorhis, and others, and involve Pearson's correction for "broad categories." The formula for biserial correlation also may be developed following these procedures. The assumptions underlying serial correlation are that the segmented variable is basically continuous and normally distributed, and that all the segments which together would form a whole normal distribution are present.

Introduction

Biserial correlation has become a standard measurement technique. Its use is indicated when the data are presented in the form of two mutually exclusive categories in respect to one variable and in quantitative scores in respect to the other variable. The assumptions underlying such a use are that the dichotomous variable is basically continuous and normally distributed, and that the two dichotomies together form a whole normal distribution.

Pearson (2, 96-105) is responsible for the first biserial expression for a correlation coefficient. Kelley (1, 248) presents the following simplified form of Pearson's biserial expression:

$$r_{bis} = \frac{(M_2 - M_1)pq}{z\,\sigma}. \qquad (1)$$

The purpose of this paper is to present a development of formulas of serial correlation for any number of categories, and more specifically formulas for biserial correlation (following a procedure other than that of Pearson or Kelley), triserial correlation, quadriserial correlation, etc. This development stems from procedures outlined by Kelley and by Peters and Van Voorhis. They used these procedures to determine correlation coefficients when the variables were segmented into several categories (1, 167-171; 4, 399-402). They did not, however, express the procedures in the form of a formula; nor did they use them in their development of the formula for biserial r (1, 245-248; 4, 362-364). This article is written principally because it would seem desirable that formulas for triserial r, etc., be explicitly presented.

The effect of serial correlation is to normalize the segmented distribution at the time that the correlation coefficient is obtained. If the number of segments is large and if the segmented variable is already normally distributed, the resulting correlation will be the same as a Pearson product-moment correlation.

Symbolism

The following symbolism will be adopted:

Let y be a continuous variable,

x be a continuous segmented variable, normally distributed, and

r be the coefficient of correlation (linear) between x and y.

Let a = the proportion of cases in the top (right-most) segment of the x distribution,

b = the proportion of cases in the second highest segment,

c = the proportion of cases in the third highest segment, etc., and

f = the proportion of cases in the f-th segment of the distribution.

Then $a + b + c + \cdots = 1$.

Let $q_a = a$,

$q_b = a + b$,

$q_c = a + b + c$, etc.,

$q_f = a + b + \cdots + f$ = the area above the left boundary of the f-th, segment, and

$q_{f-1} = a + b + \cdots +$ up to but not including f = the area above the right boundary of the f-th segment.

Let z_a = the ordinate of the normal curve, assuming a unit normal distribution, at q_a,

z_b = the ordinate of the unit normal curve at q_b, etc.,

z_f = the ordinate at q_f, and

z_{f-1} = the ordinate at q_{f-1}.

Let Y_a = the mean of the y's in the top (right-most) segment of the x distribution,

Y_b = the mean of the y's in the second highest segment, etc., and

Y_f = the mean of the y's in the f-th segment.

Let X_a = the mean of the x's in the top segment of the x distribution, etc., and

X_f = the mean of the x's in the f-th segment.

Except for a few modifications, conventional symbolism has been adhered to. The symbol q is sometimes taken in the literature to represent the proportion of cases in one of the segments (i.e., $q = q_1 - q_2$, a decumulated frequency), as well as the cumulation of proportions or frequencies from a given line of truncation to the end of the curve. In tables of the normal probability integral oriented in terms of q, and therefore in this paper, q always represents the area from the given line of truncation to the end of the curve.

Since the normal curve is symmetrical, the ordinates z are equivalent for complementary q's. Consequently most normal tables oriented in terms of q carry the argument only from zero to .500, and it is therefore necessary to consult the complement of q for values of q higher than .500. This of course does not disturb the meaning of q.

In this paper the upper segments of the x distribution (the desirable pole of the trait or measure in question) are placed to the right of the lower (or less desirable) segments of the normal curve.

General Serial Correlation

The basic formula for rectilinear correlation is

$$r = \frac{\Sigma xy}{N\sigma_x\sigma_y}. \tag{2}$$

The y variable is continuous and all of the values are known. The x variable is presented simply as an ordered sequence of categories with stated frequencies. On the assumption of a normal distribution, class-index values may be assigned each of the categories.

The means of the respective segments will be taken as the class-index values of x. The mean of a segment of a unit normal distribution, expressed as a deviation from the mean of the whole distribution, is (1, 101; 4, 290):

$$X_f = \frac{z_f - z_{f-1}}{q_f - q_{f-1}} = \frac{z_f - z_{f-1}}{f}. \tag{3}$$

In this formula the q's are the only independent variables. The values z can be determined most readily from a table of the normal probability integral oriented in terms of q.

The use of these means as class-index values representing the categories of x permits the computation of the coefficient of correlation between x and y. Pearson has shown that a coefficient so computed is in need of correction for "broad categories" (more specifically known as the correction for class means) (3, 130, as reported in 1, 167-171, and 4, 399-402). When the x variable is in terms of a

unit distribution and when no correction is necessary on account of the y variable, the corrected correlation coefficient is

$$_cr = \frac{r}{\sigma_x}. \tag{4}$$

In this equation the quantity σ_x is the *calculated* standard deviation of the index values (class means), and is less than unity, the assumed standard deviation of the true x values.

From (2) and (4), and since the distribution is of unit area,

$$_cr = \frac{\Sigma xy}{\sigma_x^2 \sigma_y}. \tag{5}$$

It may be noted that as the number of categories increases, the more nearly σ_x^2, and in turn σ_x, approach unity; and the smaller the correction necessary in the correlation coefficient.

In the case of biserial correlation, $_cr$ will be more specifically known as r_{bis}; r_{tri} in the case of triserial correlation; r_{quad} in the case of quadriserial correlation; r_{quint} in the case of quintiserial correlation; etc. More generally it will be known simply as r.

In order to find a general formula for serial correlation, it will be necessary to determine Σx^2 and Σxy. Σx can be shown to equal zero.

Now

$$\begin{aligned} \Sigma x &= aX_a + bX_b + cX_c + \cdots \\ &= \Sigma f X_f \\ &= \Sigma f \left(\frac{z_f - z_{f-1}}{f} \right) \\ &= \Sigma (z_f - z_{f-1}) \\ &= 0. \end{aligned} \tag{6}$$

$$\begin{aligned} \Sigma x^2 = \sigma_x^2 &= aX_a^2 + bX_b^2 + cX_c^2 + \cdots \\ &= \Sigma f X_f^2 \\ &= \Sigma \frac{(z_f - z_{f-1})^2}{f}. \end{aligned} \tag{7}$$

$$\begin{aligned} \Sigma xy &= aX_aY_a + bX_bY_b + cX_cY_c + \cdots \\ &= \Sigma f X_f Y_f \\ &= \Sigma Y_f (z_f - z_{f-1}). \end{aligned} \tag{8}$$

From (5), (7), and (8), it follows that:

$$r = \frac{\Sigma Y_f (z_f - z_{f-1})}{\sigma_y \Sigma \frac{(z_f - z_{f-1})^2}{f}}. \tag{9}$$

It may be of some theoretical interest to express this formula in matrix notation.

Let W and Y be column matrices:

$$\left\| \begin{matrix} z_a \\ z_b - z_a \\ z_c - z_b \\ \cdot \\ \cdot \end{matrix} \right\|, \qquad \left\| \begin{matrix} Y_a \\ Y_b \\ Y_c \\ \cdot \\ \cdot \end{matrix} \right\|,$$

and let F be a diagonal matrix:

$$\left\| \begin{matrix} \frac{1}{a} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{b} & 0 & & 0 \\ 0 & 0 & \frac{1}{c} & & \\ 0 & & & \cdot & \\ 0 & 0 & & & \cdot \end{matrix} \right\|.$$

Then

$$\Sigma Y_f (z_f - z_{f-1}) = W'Y \ ,$$

and

$$\Sigma \frac{(z_f - z_{f-1})^2}{f} = W'FW \, .$$

Both of these results are matrices of the order 1×1.

Let us indicate the scalar value of these matrices by enclosing them in braces, respectively, thus: $\{W'Y\}$; $\{W'FW\}$.

Then

$$r = \frac{\{W'Y\}}{\sigma_y \{W'FW\}}. \tag{10}$$

Biserial Correlation

Biserial r can be found by substitution in formula (9). It can also be found directly from Table 1.

TABLE 1
Calculation of Biserial r

f	q_f	q_{f-1}	z_f	z_{f-1}	x	y	fx	fx^2	fxy
a	q	0	z	0	$\frac{z}{a}$	Y_a	z	$\frac{z^2}{a}$	zY_a
b	1	q	0	z	$-\frac{z}{b}$	Y_b	$-z$	$\frac{z^2}{b}$	$-zY_b$

From the last three columns of Table 1, respectively,

$$\Sigma x = 0\,,$$

$$\Sigma x^2 = {\sigma_x}^2 = \frac{z^2}{a} + \frac{z^2}{b} = \frac{(a+b)z^2}{ab} = \frac{z^2}{ab}\,,$$

and

$$\Sigma xy = z(Y_a - Y_b)\,.$$

Therefore

$$r_{bis} = \frac{z(Y_a - Y_b)}{\frac{z^2}{ab}\sigma_y} = \frac{(Y_a - Y_b)ab}{z\,\sigma_y}\,, \tag{11}$$

which agrees with (1).

Triserial Correlation

Triserial r can be found by substitution in (9) or directly from Table 2.

TABLE 2
Calculation of Triserial r

f	q_f	q_{f-1}	z_f	z_{f-1}	x	y	fx	fx^2	fxy
a	q_a	0	z_a	0	$\frac{z_a}{a}$	Y_a	z_a	$\frac{z_a^2}{a}$	z_aY_a
b	q_b	q_c	z_b	z_a	$\frac{z_b - z_a}{b}$	Y_b	$z_b - z_a$	$\frac{(z_b - z_a)^2}{b}$	$(z_b - z_a)Y_b$
c	1	q_b	0	z_b	$-\frac{z_b}{c}$	Y_c	$-z_b$	$\frac{z_b^2}{c}$	$-z_bY_c$

From the last three columns of Table 2, respectively,

$$\Sigma x = 0\,,$$

$$\Sigma x^2 = {\sigma_x}^2 = \frac{z_a^2}{a} + \frac{(z_b - z_a)^2}{b} + \frac{z_b^2}{c}\,,$$

and

$$\Sigma xy = z_a Y_a + (z_b - z_a) Y_b - z_b Y_c .$$

Therefore

$$r_{tri} = \frac{z_a Y_a + (z_b - z_a) Y_b - z_b Y_c}{\sigma_y \left[\frac{z_a^2}{a} + \frac{(z_b - z_a)^2}{b} + \frac{z_b^2}{c} \right]} . \tag{12}$$

Quadriserial Correlation

It can be shown similarly that

$$r_{quad} = \frac{z_a Y_a + (z_b - z_a) Y_b + (z_c - z_b) Y_c - z_c Y_d}{\sigma_y \left[\frac{z_a^2}{a} + \frac{(z_b - z_a)^2}{b} + \frac{(z_c - z_b)^2}{c} + \frac{z_c^2}{d} \right]} . \tag{13}$$

Quintiserial Correlation

It can also be shown that

$$r_{quint} = \frac{z_a Y_a + (z_b - z_a) Y_b + (z_c - z_b) Y_c + (z_d - z_c) Y_d - z_d Y_e}{\sigma_y \left[\frac{z_a^2}{a} + \frac{(z_b - z_a)^2}{b} + \frac{(z_c - z_b)^2}{c} + \frac{(z_d - z_c)^2}{d} + \frac{z_d^2}{e} \right]} . \tag{14}$$

Standard Error of Serial Correlation

In 1914, several years after Pearson published his formula for biserial correlation, Soper published formulas for the standard error of biserial correlation (1, 248-249; 4, 365). Formulas for the standard error of serial correlation in general have not yet been developed.

Illustration

Test scores are available for a group of 40 workers, of whom a foreman rates 10 outstanding, 24 adequate, and 6 weak.

The test scores of the workers rated outstanding are: 115, 112, 108, 104, 103, 100, 95, 92, 86, 85.

The test scores of the workers rated adequate are: 110, 108, 106, 105, 101, 100, 98, 95, 95, 92, 90, 90, 89, 88, 87, 87, 85, 84, 79, 78, 76, 74, 73, 70.

The test scores of the workers rated weak are: 90, 87, 82, 78, 73, 70.

The following statistics can be computed immediately from the test score data: $Y_a = 100$; $Y_b = 90$; $Y_c = 80$; $\sigma_y = 12.1120$; $a = .25$; $b = .60$; $c = .15$.

A table such as Table 3 can be set up to find the other constants of formula (12).

TABLE 3
Normal Curve Functions

	f	q	z
a	.25	.25	.3178
b	.60	.85	.2332
c	.15	1.00	0

By substitution directly in (12),

$$r_{tri} = \frac{(.3178)(100) + (.2332 - .3178)(90) - (.2332)(80)}{12.1120\left[\frac{.3178^2}{.25} + \frac{(.2332 - .3178)^2}{.60} + \frac{.2332^2}{.15}\right]}$$

$$= .584 .$$

The obtained correlation should not be accepted as the validity of the test without further inquiry into such important factors as the validity of the criterion, the rectilinearity of regression, and the heterogeneity of the sample relative to that of the applicant population for whom the test is intended.

REFERENCES

1. Kelley, T. L. Statistical method. New York: MacMillan, 1923.
2. Pearson, K. On a new method of determining correlation between a measured character A and a character B, of which only the percentage of cases wherein B exceeds or falls short of a given intensity is recorded for each grade of A. *Biometrika*, 1909, 7, 96-105.
3. Pearson, K. On the measurement of the influence of "Broad Categories" on correlation. *Biometrika*, 1913, 9, 116-139.
4. Peters, C. C. and Van Voorhis, W. R. Statistical procedures and their mathematical bases. New York: McGraw-Hill, 1940.

PSYCHOMETRIKA—VOL. 11, NO. 1
MARCH, 1946

A FACTORIAL ANALYSIS OF OLFACTION

E. H. HSÜ*
THE UNIVERSITY OF CHICAGO

Twenty-one reagents were rated on a pleasantness scale by 182 subjects. The tetrachoric correlation coefficients were determined for the variables. The correlation matrix thus obtained was subjected to a Thurstone analysis which gave rise to a 6-dimensional structure. Two of these were not clear-cut, and the last one appeared to be a residual factor. One factor, the benzene-ketone, revealed the chemical property of unsaturation and the physiological quality of possible trigeminal stimulation. The other two factors were marked by the presence of oxygen and nitrogen, respectively, and were tentatively labelled as a plant factor and an animal factor accordingly. The intercorrelations of the primaries are high.

Introduction

There seems to be some justification in classifying the tactual, the auditory, and to some extent the visual sensations under the heading of physical sense and the olfactory and gustatory sensations under that of chemical sense. Possibly because of the overgrowth of the corpus callosum in the human brain the growth of the archipallium is necessarily sacrificed. This may possibly explain why little is known about the more primitive olfactory sensation in human beings. For instance, such attempts at searching for the primary dimensions of olfaction as carried out by Zwaardemaker and Henning have never been critically accepted, and we still know almost nothing about the primary olfactory sensations, and still less about the differential olfactory receptors.

The present study aims at singling out the primary factors of olfaction by resorting to the modern factorial technique, which may at least throw light on the number of dimensions of primary olfaction. Unfortunately, a research on such a scale calls for a large sampling of olfactory stimuli which are unavailable for our purpose in war time. It should be pointed out that since our problem is essen-

* The author wishes to express his gratitude to Professor L. L. Thurstone for his kindness in providing facilities which made this study possible and for his guidance on the factorial problem; to Doctor T. G. Andrews for his many invaluable assistances and technical advice on the experimental aspect of this problem; to Miss Florence Brown for her editorial assistance; and to Mrs. Virginia Brown for her generous assistance on the computational procedures.

tially a physiological one more objective measures than psychological ratings are desirable. Perhaps such measures as the adaptation rate, threshold values, and the like are the ideal variables for isolating the primary olfactory factors. To carry out such experiments on a large group of individuals with a large number of stimuli would call for a considerable length of time, however. Time at our disposal did not allow us to go into such elaborated experiments. We therefore resorted to attacking this problem by making use of the psychological reaction towards a relatively small sample of stimuli as a good preliminary experiment.

An earlier factorial study on the olfactory problem was carried out by Beebe-Center (1). Using Spearman's method of tetrad differences he found a general factor from the correlations of rankings of fourteen odors by eight subjects. Due to such a small sample and his factorial approach, the result was not significant. A more recent work of Eysenck (8) made a reanalysis of the Beebe-Center data. In addition to the reanalysis Eysenck used fourteen subjects, eight men and six women, to arrange thirty-one smells according to a scale of predescribed frequencies for scores from 0 to 10. The stimuli were first sorted into three groups, namely, the definitely pleasant, the definitely unpleasant, and the indifferent. Then a score of 10 was assigned to the best-liked odor and a score of 9 to the next two best and so on. Burt's least squares method was applied to the resulting correlation matrix between persons, and factor saturations were obtained. A general factor was found, while the residual coefficients were not significant. Eysenck attributed this general factor to biological reasons. She also presented the average ratings for each grouping of Zwaardemaker and believed the hedonic tones thus expressed were in agreement with his classification.

Some less recent works connected with the olfactory problem are briefly reviewed here. A series of works were conducted by Elsberg and his associates (2, 3, 4, 5, 6, 7). Two new methods of quantitative measurement of olfaction were invented. One was called the "blast injection" method. In this method air is injected into the patient's nostrils, either monorhinally or birhinally, through a rubber tubing and a nosepiece by means of a tank of compressed air. The bottle containing the reagent is connected with the tank by rubber tubing at one end and with the nosepiece at the other. Thus the volume of air injected into the patient's nasal passages can be measured. The intensity of the odor is thus controlled by allowing a definite amount of air to pass through the mucous membrane of the nostrils. The patient is asked to hold his breath while the air is injected. By this method Elsberg determined the threshold values of some twenty-

one odors. The other technique was the so-called "stream injection" method. This was devised for determining olfactory fatigue. In this method a continuous stream of air is injected into the patient's nostrils until the patient can no longer notice the odor. While the air is being injected he is asked to breathe through his mouth. Elsberg obtained from patients the olfactory fatigue data for coffee by a birhinally injected stream of coffee odor for 1.5 to 2.0 minutes at a volume rate of 6000 cc to the minute. He found that the olfactory fatigue of coffee did not interfere with the ability of the individual to reproduce by an effort of memory the sensation of that odor; however the reproduction necessitated a greater effort than that normally demanded. Elsberg assigned the identification of odor to the olfactory cortex and reception and perception to the thalamus or to its neighboring area. He also found that if the olfactory membrane was stimulated by an odor at the moment when the individual was concentrating on the same odor a smaller volume of odor than that normally required could be identified.

The Procedure of the Experiment

A preliminary list of thirty chemical reagents was made according to the following criteria: (1) They had to be relatively free from trigeminal stimulation as based on Von Skramlik's findings (12). (2) They had to be chemically pure. (3) They had to represent a large variety of olfactory quality. Only twenty-two reagents, however, were available. One of these, mercaptan, was finally discarded due to its extremely offending nature. The list of the twenty-one reagents finally used in this experiment is as follows.

1. 0-Eugenol 2. Benzaldehyde 3. α-Ionone 4. Diethyl sulfide 5. Indol 6. Naphthalene 7. Eucalyptol 8. Anethol 9. Acetone 10. Trimethyl amine 11. α-Pinene 12. Geraniol 13. α-Terpineol 14. Skatol 15. 0-Xylol 16. Linaloöl 17. Toluene 18. Amyl acetate 19. Methyl salicylate 20. Limonene 21. Pyridine

These twenty-one reagents were then bottled in identical 500-cc containers, each with a glass stopper. The quantities of the liquid reagents varied from 2 cc (trimethyl amine) to 10 cc. The majority were 10 cc. Diethyl sulfide and pyridine were about 5 cc each, and indol and skatol were about 5 grams.

With the exception of a few cases in which the experiment was conducted individually, the odors were given to the majority as a group experiment. The number of individuals in each group varied

from two to twenty-five. In most cases the group was small, with about four persons at a time. The following data sheet was given each subject:

Name (optional): Age: Sex: MF
Check the following:
How many cigarettes do you smoke a day? 0 1-5 6-10 11-15 16-20 Over
Have you smoked within one hour before this experiment? Yes No
Do you have a cold? No Slightly Definitely
Instruction:
You will be handed a stoppered bottle on which there is a *number*. At a signal from the experimenter you are to uncork the bottle, smell the cork and immediately restopper the bottle. As you smell the cork, *pay particular attention to the relative pleasantness or unpleasantness of the odor*. On the graded blank given you, identify the degree of pleasantness at which the odor was judged by you. Mark the scale *for the number of the bottle smelled, in the proper column*.

At the next signal from the experimenter pass the bottle to the next subject and receive the next numbered bottle, waiting for the signal to unstopper and smell, and evaluate as described above.

It is of utmost importance that you take this experiment seriously and attempt to give your best judgments concerning the odors. For your cooperation as subjects in this research project the experimenter and the staff of the Department of Psychology are extremely grateful.

Reagent No.	Unpleasant	Slightly unpleasant	Indifferent	Slightly pleasant	Pleasant
1					
2					
3					

A section of the graded blank which covered the twenty-one reagents is shown above. In addition to this written instruction the following oral instruction was given all subjects: "When you open the bottle, you are to open it little by little until you smell a distinct quality of odor. But be sure you smell a distinct odor, not any diffused odor attached to the outside of the bottle. If you cannot get a distinct odor, then smell the cork. If you have two different impressions of the same odor, then take your first impression as the basis for judging the quality of the odor. Always sniff just slightly until you get the distinct quality of the odor. Don't smell deeply any odor. If you cease to smell any odor at all, please tell me and take a little rest." This additional instruction was given in order to avoid adaptation and to keep the laboratory free from odors leaking too much from the bottles while they were being opened. The laboratory windows were always opened in such a way that good ventilation was assured. The

bottles containing the odors were put far from the subjects before and after the main procedure. The numbered bottles were given entirely in random order, so as to eliminate the constant error of contrast of the hedonic tone. When the group was small, the experimenter uncorked the bottles for each subject in order to avoid adaptation as much as possible and to give a relatively lengthy interval for each subject between smellings. It took about ten minutes on the average for each subject to finish the whole series of odors.

The subjects included in this study consisted of 182 individuals, 71 males and 111 females. With a few exceptions the subjects were students, the majority of them taking courses in psychology. The ages of the subjects ranged from 15 to 53 years, with a median of 23. Of 182 subjects 102 reported being non-smokers, 76 reported having slight colds and 48 reported having smoked within one hour of the experiment.

The Result

A brief analysis of the factors of sex, smoking habits, and colds did not reveal any significant difference as far as the ratings of the olfactory stimuli were concerned. Since the average age of the subjects was about 22 and the variation was small, the age factor seemed also insignificant. The frequencies of ratings of each reagent on the 5-point scale were obtained first, and then the scale was dichotomized near the median. Since the scale was considered linear, by such dichotomizing a four-fold tetrachoric table for each paired variable could easily be determined. Then the tetrachoric correlation coefficients were obtained according to Thurstone's diagrams (11). The correlation matrix so determined is presented in Table 1.

The correlation matrix was then subjected to a centroid analysis with Professor Thurstone's new method (9). After the 6th factor had been extracted the residuals appeared negligible. None of the residuals was greater than .20, the standard error for a zero tetrachoric correlation in our case being about .21. The centroid matrix is given in Table 2. The residuals are shown in the upper half above the diagonal in Table 1.

The rotation was carried out six times for all planes with the method of real length of vectors on a matrix multiplier machine. Plane A was then rotated with Professor Thurstone's newly invented rotation method of least squares three times (10), Planes C and B four times, Plane E three times, and Plane F one time. It was found that the new rotation method proved to be the most powerful technique yet discovered. In comparison with the older methods it saved a considerable number of rotations. The transformation matrix is

presented in Table 3. The simple structure is given in Table 4.

From Table 5 it can be seen that the primaries are highly correlated. There possibly exists some second-order factor.

Interpretation of Result

First it should be pointed out that the nature of rating in our experiment is essentially different from hedonic ratings on visual or auditory stimuli in aesthetic experiments. All of the previous workers, such as Beebe-Center and Eysenck, seem to have treated olfactory ratings and visual and auditory aesthetic ratings alike. They appear to us to differ in three ways: (1) In an aesthetic rating one may measure or describe some more fundamental quality in vision or audition, such as hue or pitch, besides a pleasantness or unpleasantness rating. But there is no other quality of olfaction more fundamental than pleasantness or unpleasantness. (2) The aesthetic ratings on visual or auditory stimuli do not ordinarily give rise to such physiological reaction as nausea or headache, but the odors do if they are unpleasant. (3) The reaction to odors is more spontaneous than that towards visual or auditory stimuli. In other words the variance of aesthetic rating on visual or auditory stimuli is larger than that of olfactory rating. For example, out of a number of lines one may either choose a line of three inches or one of two and a half as his preferred one, but he is almost bound to be offended by such an odor as mercaptan, if his olfactory receptors are not too different from those of the majority of people. Thus it seems quite logical for us to use the pleasantness-unpleasantness ratings on the olfactory stimuli to search for their more fundamental qualities and primary dimensions.

It is to be noted in Table 6 that a positive manifold definitely exists in our simple structure. This fact leads us to believe that the olfactory correlation matrix resembles the correlation matrix of scores on intelligence tests more than do correlation matrices of personality traits, in which a positive manifold in general cannot be demonstrated.

Factor A reveals these following variables:

Acetone	.75	Toluene	.57	Diethyl sulfide	.40
0-Xylol	.73	Naphthalene	.51		

Amyl acetate and indol also appeared in this group but had relatively low loadings. An inspection of these variables would reveal only two derivatives, namely benzene and ketone, with the exception of diethyl sulfide. The appearance of the thioether is probably due to some error made in the experimental procedure under which the subjects

might not have really smelled the quality of this chemical but got some diffused odor attached to the outside of the bottle instead. Nevertheless, it should be pointed out that all these reagents with high loadings have something definitely in common psychologically, since they all give rise to a biting sensation, which is possibly due to the stimulation of the trigeminal nerve. Chemically, benzene and ketone are not alike, yet it is interesting to note that in this factor all the reagents have double-bond in their structures; hence all are unsaturated. The aromatic hydrocarbons are also grouped together by their loadings, which reveal their chemical resemblance.

Factor B consists of the following variables:

Linaloöl	.50	Methyl salicylate	.43	Ionone	.31
Benzaldehyde	.49	O-Eugenol	.43		
Anethol	.47	Geraniol	.35		
Terpineol	.46	Amyl acetate	.35		

The striking feature of this factor is the presence of oxygen in all the reagents. Half of them are alicyclic compounds and practically all are constituents of plants. If one examines the structures of these chemicals one sees a single oxygen atom either attached to the carbon, or in a hydroxyl group, or combined with the methyl group. The oxygen common to all these reagents is responsible for the pleasant odor of this group. Thus this factor is both chemically and psychologically meaningful. We tentatively label this factor as a plant factor.

The following variables are found in Factor C:

Toluene	.54	Diethyl sulfide	.36
Pinene	.51	Pyridine	.33
Linaloöl	.48	Terpineol	.26
Limonene	.44	Geraniol	.25

This factor is less clear-cut than the others. Possibly the last two reagents in this factor are not significant, since their loadings are low. The double-bond in their structures seems to be the only chemically common nature of these reagents. This factor thus consists of two groups, benzene (if pyridine is considered a derivative of benzene) and alicyclic compounds which appear to be a combination of Factor A and Factor B.

Factor D includes the following variables:

O-Eugenol	.42	Eucalyptol	.34
Pinene	.40	Trimethyl amine	.33
Methyl salicylate	.40	Limonene	.30
Skatol	.36	Indol	.24

The variables in this factor can be dichotomized chiefly into the

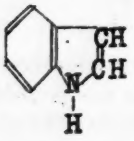
alicyclic and heterocyclic compounds. It is interesting to note that nitrogen is present in the heterocyclic compounds and in trimethyl amine. Nitrogen is connected with protein, which is characteristic chiefly of animals. This factor seems to indicate a combination of plant and animal.

The following variables are found in Factor E:

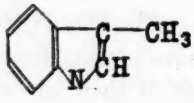
Indol	.62	Eucalyptol	.48
Skatol	.60	Trimethyl amine	.36
Naphthalene	.49	Amyl acetate	.32
Benzaldehyde	.38	Methyl salicylate	.28

The structures of indol and skatol are given as follows:

Indol



Skatol



It is to be noted that the three compounds possessing nitrogen again appear in this factor. The resemblance of indol and skatol, chemically, physiologically, and psychologically, is revealed by their remarkably high loading on this factor. We have included only four reagents possessing nitrogen—indol, skatol, trimethyl amine, and pyridine, all of which are directly or indirectly related to animal constitution. The fact that pyridine or bone oil did not appear in this factor clearly indicates the possibility of grouping the other three together. It is highly possible that an exclusive factor of composed reagents possessing nitrogen would have appeared had we only included more such compounds in our sample. However, in the present case it seems safe to label this factor as an animal one because of the two highest loadings on this factor.

The following reagents appeared in Factor F:

O-Eugenol	.57	Benzaldehyde	.28
Methyl salicylate	.37	Ionone	.26
O-Xylol	.36	Acetone	.26
Trimethyl amine	.34	Anethol	.25

There is only one high loading on this factor. The picture of this factor is not clear-cut, and it seems to have all the preceding factors combined in it. Thus it is not interpretable and is most likely just a residual factor.

It should be pointed out that a few of our subjects whose nasal passages had been operated on for one reason or another were either totally insensitive to some reagents, for instance, indol, or had a marked deviation from the majority rating of degree of pleasantness. In further studies it would be interesting to score such individuals according to the regression weights derived from the factors.

TABLE 1
Correlation Matrix (Each figure is to be multiplied by .01. Above the diagonal cells are the residuals with the sixth factor removed.).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1		-04	-05	06	-01	-07	-01	-06	-04	01	-01	-15	07	00	01	09	05	06	10	-02	04
2	05		01	-02	-04	02	01	-01	00	00	00	06	-08	-01	03	-07	00	11	-04	10	01
3	-01	21		-04	04	-05	-01	02	-04	-03	06	-02	04	02	10	04	-03	00	00	02	-04
4	-10	05	-01		-14	00	03	12	-06	-02	06	-16	-03	02	-11	-05	05	09	-02	00	09
5	-05	20	00	01		04	02	11	-01	-12	00	16	-09	09	01	-02	-05	-12	-07	03	16
6	-06	20	-17	20	44		-04	-04	04	08	-05	-01	02	-08	05	-06	02	03	05	13	-01
7	17	07	-22	-06	26	29		08	09	13	09	-12	11	-09	-03	08	-06	02	-07	-02	-11
8	25	16	10	13	11	10	21		00	05	01	11	-13	-09	-15	-04	-04	-04	-03	-03	07
9	12	04	-09	13	18	41	10	21		06	-09	01	06	05	09	11	-05	-12	-07	01	-09
10	02	18	08	-02	13	08	12	-07	-01		10	-02	05	06	-07	05	00	-15	-01	-07	02
11	05	04	02	22	15	10	25	05	-15	17		-11	-09	-14	09	02	06	10	-09	03	00
12	-01	14	04	-08	07	06	-10	37	08	-23	-03		06	09	08	-06	-03	-10	00	15	-04
13	07	18	19	09	-10	08	16	04	-11	-01	05	31		-01	02	07	05	-11	00	-10	-05
14	-12	18	-02	20	53	20	10	-21	05	38	13	-05	03		04	-04	06	05	07	01	-07
15	10	10	12	11	29	38	-12	-05	74	04	05	03	-20	16		-03	00	-04	-02	-01	-06
16	27	13	17	10	-05	02	13	24	10	-05	26	27	45	-05	-13		02	-01	-09	-03	00
17	10	01	-03	38	10	29	-10	08	32	-03	34	12	10	23	37	25		03	14	-17	01
18	10	38	04	25	12	40	20	16	13	-16	17	08	10	20	16	20	20		13	-12	07
19	60	12	-05	-12	07	32	32	33	20	01	02	18	05	09	11	10	22	35		01	04
20	14	08	-05	12	08	25	10	10	10	-12	30	30	-01	10	03	20	10	-07	20		05
21	-14	16	16	34	27	-05	-38	-04	-12	27	14	-09	05	17	09	09	21	08	-24	09	

1. O-Eugenol 2. Benzaldehyde 3. α-Ionone 4. Diethyl sulfide 5. Indol
6. Naphthalene 7. Eucalyptol 8. Anethol 9. Acetone 10. Trimethyl amine
11. α-Pinene 12. Geraniol 13. α-Terpineol 14. Skatol 15. O-Xylol 16. Linalool 17. Toluene 18. Amyl acetate 19. Methyl salicylate 20. Limonene
21. Pyridine

TABLE 2
Centroid Matrix

	I	*II*	*III*	*IV*	*V*	*VI*	h^2
1	.29	.40	—.24	—.21	—.28	—.31	.52
2	.42	.14	.11	.06	.40	—.18	.40
3	.09	.11	.26	.18	.12	—.31	.28
4	.32	—.27	.21	.23	—.06	.15	.30
5	.45	—.33	—.14	—.19	.32	.07	.47
6	.54	—.17	—.31	.05	.16	.31	.54
7	.26	.19	—.33	—.40	.13	.31	.49
8	.36	.38	—.13	.13	—.12	—.07	.33
9	.43	—.24	—.49	.38	—.23	—.09	.69
10	.16	—.27	.12	—.30	.25	—.35	.39
11	.36	—.06	.27	—.30	—.18	.21	.37
12	.25	.36	.10	.22	—.14	.14	.29
13	.27	.35	.37	.13	.19	.17	.41
14	.38	—.39	.13	—.35	.27	.11	.52
15	.42	—.48	—.36	.32	—.14	—.25	.72
16	.43	.38	.33	.10	—.13	.12	.48
17	.49	—.27	.10	.17	—.37	.14	.51
18	.47	.09	—.11	.21	.23	.15	.36
19	.47	.40	—.43	—.22	—.11	—.07	.63
20	.32	.05	.10	—.14	—.33	.15	.27
21	.20	—.35	.48	.09	.06	—.20	.44

TABLE 3
Transformation Matrix, A

	A	*B*	*C*	*D*	*E*	*F*
I	.53	.51	.46	.38	.52	.39
II	—.48	.74	—.08	—.01	—.16	.10
III	—.32	.12	.63	.01	—.17	—.11
IV	.54	.21	—.05	—.86	—.48	—.11
V	—.25	.24	—.55	—.26	.65	—.10
VI	.18	—.29	.28	—.21	.15	—.89

TABLE 4
Simple Structure

		A	*B*	*C*	*D*	*E*	*F*
1.	O-Eugenol	—06	39	03	42	00	51
2.	Benzaldehyde	02	49	—02	04	38	28
3.	Ionone	—08	31	03	—08	—07	26
4.	Diethyl sulfide	40	—02	36	—08	05	—08
5.	Indol	27	—01	00	24	62	08
6.	Naphthalene	51	07	06	05	49	—07
7.	Eucalyptol	—04	09	—07	34	48	—09
8.	Anethol	14	47	09	07	00	25
9.	Acetone	75	03	—01	—09	00	26
10.	Trimethyl amine	—11	00	—05	33	36	34
11.	Pinene	05	00	51	40	21	—03
12.	Geraniol	11	38	25	—09	—12	—01
13.	Terpineol	—09	46	26	—09	11	—08
14.	Skatol	11	—12	19	36	60	01
15.	O-Xylol	73	—08	00	—02	07	36
16.	Linaloöl	05	50	48	09	—01	06
17.	Toluene	57	—03	54	11	—02	05
18.	Amyl acetate	32	35	04	—09	32	02
19.	Methyl salicylate	09	43	—03	40	28	37
20.	Limonene	15	06	44	30	02	03
21.	Pyridine	12	—01	33	03	04	15

TABLE 5
Cosines of Angular Separations Between Reference Vectors ($\Lambda'\Lambda$) and Primary Vectors ($D(\Lambda'\Lambda)^{-1}D$) (Correlations between reference vectors are below the diagonal cells.)

	A	*B*	*C*	*D*	*E*	*F*
A'	1.00	.51	—.60	.64	—.50	—.55
B'	—.13	1.00	—.52	.62	—.52	—.72
C'	.24	.03	1.00	—.73	.62	.61
D'	—.23	.00	.31	1.00	—.72	—.77
E'	.01	.14	—.15	.41	1.00	.56
F'	.00	.47	—.07	.45	.06	.99

REFERENCES

1. Beebe-Center, J. G. Pleasantness and unpleasantness. New York: Century, 1933.
2. Elsberg, C. A. and Levy, I. The sense of smell. 1. A new and simple method of quantitative olfactometry. *Bull. Neurol. Institute of New York,* 1935, **4,** 5-19.
3. Elsberg, C. A. and Brewer, E. D. The sense of smell. II. A new principle for the classification of odors based upon their olfactory coefficients. *Bull. Neurol. Institute of New York,* 1935, **4,** 20-25.
4. Elsberg, C. A. The sense of smell. VIII. Olfactory fatigue. *Bull. Neurol. Institute of New York,* 1935, 4, 479-499.
5. Elsberg, C. A. and Brewer, E. D. The sense of smell. X. A detailed description of the techniques of two olfactory tests used for the localization of supratentorial tumor of the brain. *Bull. Neurol. Institute of New York,* 1935, **4,** 501-510.
6. Elsberg, C. A. The sense of smell. XIV. The relation of the cerebral cortex to the olfactory impulse and the area of brain involved in fatigue of the sense of smell. *Bull. Neurol. Institute of New York,* 1937, **6,** 118-125.
7. Elsberg, C. A. and Stewart, J. Quantitative olfactory tests. *Archives Neurol. and Psychiatry,* 1938, **40,** 471-481.
8. Eysenck, M. D. An experimental and statistical study of olfactory preference. *J. exp. Psychol.,* 1944, **34,** 246.
9. Thurstone, L. L. Multiple factor analysis. (In press)
10. Thurstone, L. L. A single plane method of rotation. (In press)
11. Thurstone, L. L. et al. Computing diagrams for tetrachoric correlation coefficients. Univ. of Chicago Bookstore, 1935.
12. Woodworth, R. S. Experimental psychology. New York: Holt, 1938.

PSYCHOMETRIKA—VOL. 11, NO. 1
MARCH, 1946

GRAPHICAL AND TABULAR AIDS FOR DETERMINING SAMPLE SIZE WHEN PLANNING EXPERIMENTS WHICH INVOLVE COMPARISONS OF PERCENTAGES

FRANCES SWINEFORD
THE UNIVERSITY OF CHICAGO

When one knows in advance the minimum percentage difference which will be of practical interest in an investigation, principles of economy suggest the desirability of an accurate estimate of the size of the sample to be used. Charts are presented which eliminate any computation where the two groups involved are to be equal. A table is included for the case where one group will be from one to three times the size of the other. The charts are also useful for determining whether or not obtained differences are statistically significant at the 5 per cent or the 1 per cent level.

In planning an experimental investigation it is desirable to determine in advance the number of cases which are necessary in order that certain results shall be considered statistically significant. An underestimate leads to inconclusive results; an overestimate may entail needless labor and expense. Such a problem recently came to the writer's attention. In a particular investigation percentage differences of ten or fifteen points were considered meaningful; that is, in this instance only differences of at least ten or fifteen points would be of practical value, and the use of groups large enough to establish the statistical reliability of differences smaller than these would not be necessary. For how large a sample are differences of this magnitude significant at the 5 per cent level? at the 1 per cent level? If the nature of the data is such that all percentages tend to be very high or very low, then how much smaller may the groups be if they are still to yield statistically reliable results?

Examination of published tables and graphs which short-cut computations for comparing frequencies or percentages failed to reveal a ready answer to this particular problem.* Most of the graphs are designed for the experimenter who has his data collected and his number of cases, therefore, fixed. A change in the number of cases neces-

* For a comprehensive list of such references see Donald W. Fiske and Jack W. Dunlap. A graphical test for the significance of differences between frequencies from different samples. *Psychometrika*, 1945, 10, 225-29.

sitates a new graph. The advantage of the charts to be presented here lies in the fact that the number of cases, as well as the percentages and the differences, is a variable. The fixed values are the critical ratios, which have been selected at the 5 per cent and the 1 per cent points.

The discussion which follows is written in terms of proportion instead of percentage. If, as is most frequently the case, the null hypothesis to be tested is that no population difference exists, then the appropriate estimate of the population proportion is

$$p = \frac{N_1 p_1 + N_2 p_2}{N_1 + N_2}, \tag{1}$$

where p without a subscript is the weighted average of the proportions, p_1 and p_2, for the two groups, and N_1 and N_2 are the numbers of cases involved. The square of the standard error of either observed proportion is pq (where $q=1-p$) divided by the corresponding number of cases, and the standard error of the difference ($d=p_1-p_2$) is found by means of the formula,

$$\overline{SE}_d^2 = pq \frac{N_1 + N_2}{N_1 N_2}. \tag{2}$$

The critical ratio is computed from

$$\overline{CR}^2 = \frac{N_1 N_2 d^2}{(N_1 + N_2) pq}. \tag{3}$$

If N_2/N_1 be denoted by k, then formula (3) may be written

$$\overline{CR}^2 = \frac{k N_1 d^2}{(1 + k) pq}, \tag{4}$$

from which

$$N_1 = \frac{(1 + k) pq \, \overline{CR}^2}{k d^2}. \tag{5}$$

The total number of cases is then given by

$$N_1 + N_2 = (1 + k) N_1 = \frac{(1 + k)^2 pq \, \overline{CR}^2}{k d^2}. \tag{6}$$

Many problems, such as those dealing with test-item evaluation, involve equal groups, in which event k is equal to 1, and formula (5) reduces to

$$N = \frac{2\,pq\,\overline{CR}^2}{d^2}. \qquad (5a)$$

For a given critical ratio and a given difference, d, the value of N reaches a maximum when $p = q = .50$, and becomes smaller as p approaches 0.00 or 1.00. Accordingly, values of N have been computed for values of pq corresponding to $p_1 = .08$ to $p_1 = 1.00$, where p_1 represents the larger proportion, and for values of d from .08 to .30. This has been done at the 5 per cent level ($CR = 1.960$) and at the 1 per cent level ($CR = 2.576$). The results are presented in graphical form in Charts 1 and 2. Thus, for example, from Chart 1 it can readily be seen that any difference of 10 per cent is statistically significant at the 5 per cent level only if there are at least 192 cases in each group. From Chart 2, on the other hand, it is found that for 192 cases in each group a difference of 10 per cent is significant at the 1 per cent level only if the larger proportion is about .23 or smaller or about .87 or larger. Otherwise up to 332 cases would be required in each group.

The charts can also be used to determine whether or not an observed difference is significant at either of the critical levels for which the charts were drawn. Usually one wishes to know only the relation of the observed data to the 5 per cent or the 1 per cent point; the exact amount of deviation from the critical point is less important. Suppose, for instance, one has the following data: $N_1 = N_2 = 60$, $p_1 = .35$, $p_2 = .18$. The point on Chart 1 for $p_1 = .35$ and $N = 60$ is approximately on the curve for $d = .16$. The observed d of .17, therefore, is statistically significant at the 5 per cent level. The corresponding point on Chart 2, however, is very near the curve for $d = .20$. Hence the observed difference does not reach the 1 per cent level of significance. (The actual critical ratio for this example is $17/.0806 = 2.11$.)

The problem is somewhat less simple when the two groups differ in size. Minimum values of N_1 and N_2 have been computed for selected values of k and d by employing formula (5) with $pq = .25$ and with $CR = 1.960$ for the 5 per cent points and $CR = 2.576$ for the 1 per cent points. These values, given in Table 1, correspond to the maximum points on curves similar to those in the charts. Indeed, the last line of Table 1 gives the maximum points on the curves in the charts.

In order to make use of the table it is assumed that the investigator knows approximately the relative sizes of the groups to be compared and the minimum percentage difference which would be of practical import. For example, he may be sampling a college population in which there are about twice as many freshmen as seniors, and

he may wish to obtain a sufficiently large sample that a difference of .20 would be statistically significant at the 1 per cent level. Since one group is twice as large as the other, he would enter the table in the row corresponding to $k = 2.0$, and he would find the bold face (1 per cent) values of N_1 and N_2 in the column for $d = .20$. These values are 62 and 124. He will therefore require 186 cases, of which 124 are freshmen and 62 are seniors. It is not so important that he use these exact numbers as it is that he take care not to make an unduly large under- or overestimate in planning his investigation.

If it should be known in advance that the observed proportions will lie near one of the extremes, so that pq will not exceed some fraction of its maximum value of .25, say $pq/.25 \leq m$, then the minimum sample necessary for significant results is m times that entry in Table 1 which corresponds to the k and the d under consideration. For example, suppose the items on a particular questionnaire to be consistently answered in a certain way by 80 to 100 per cent of the respondents. Suppose further that a percentage difference of 8 or more is an important difference, and that one of the groups under comparison is 1.5 times as large as the other. For $k = 1.5$ and $d = .08$, Table 1 indicates that a total group of $432 + 648 = 1080$ cases would be required for significance at the 1 per cent level when $pq = .25$. In the present example, however, the maximum pq will be only .16, from which m is .64. The product, (.64) (1080), gives 691 cases as the total number required for the experiment. There should then be 276 cases in one group and 415 cases in the other.

With regard to interpolated values from the table, it should be pointed out that the only linear relation is that for successive N_2's in a column. Since N_1 changes slowly, however, vertical linear interpolation for N_1 will not introduce much error. This value can also be determined from the expression, $N_1 = N_2/k$. Horizontal linear interpolation will tend to overestimate the N's by a small number of cases.

Both the charts and the table should serve to emphasize the importance of an adequate sample in any investigation which involves comparison of percentages. It is obvious that, unless one is prepared to ignore differences of less than 20 percentage points, a total group of more than 100 cases is necessary for significance at the 5 per cent level and one of more than 200 cases, at the 1 per cent level. In the event that smaller differences have meaning for the experimenter, he can establish their existence beyond reasonable doubt only by employing very large samples.

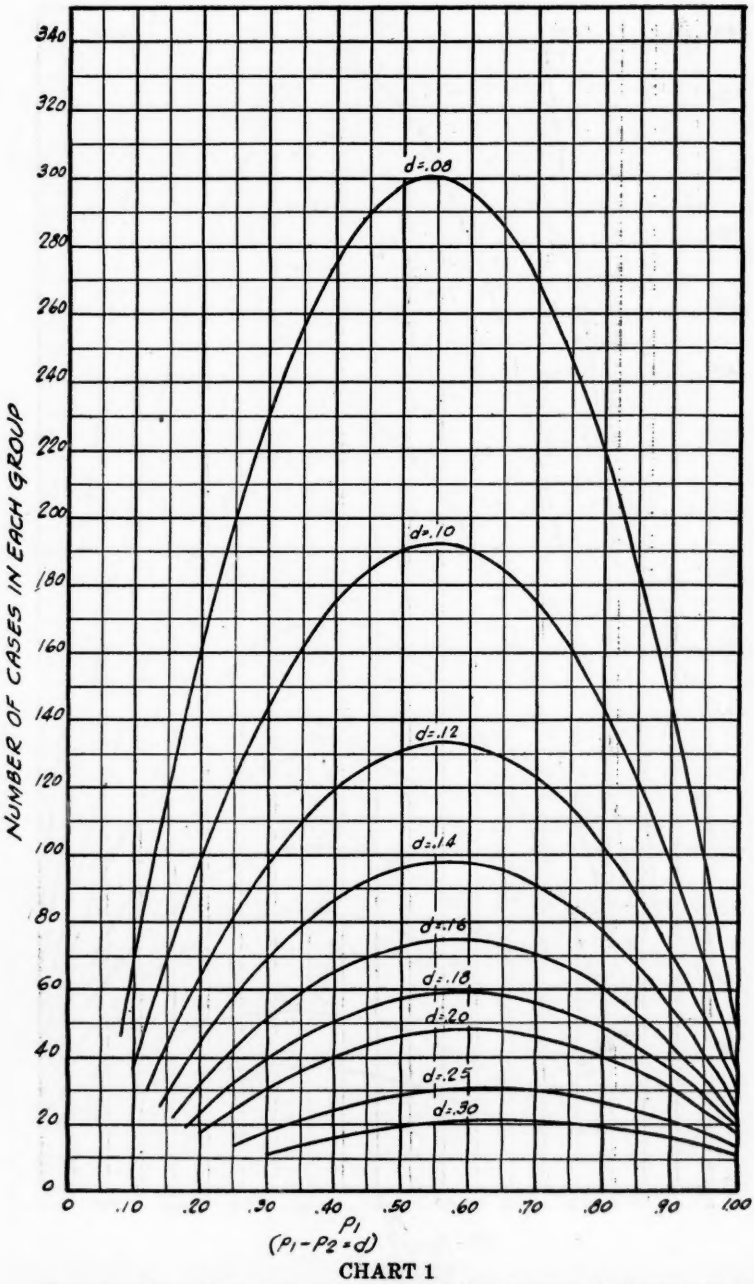


CHART 1

Corresponding Values of N and p_1 for Selected Values of d (5 per cent level of significance).

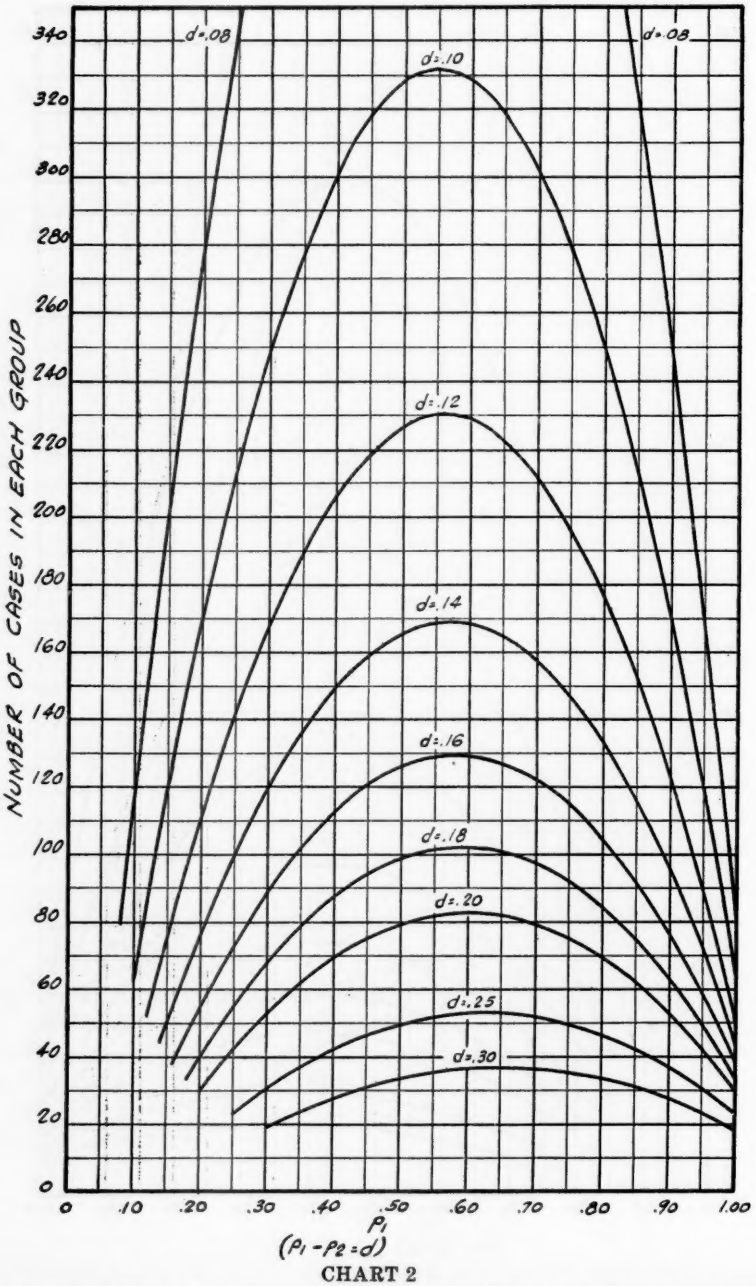


CHART 2

Corresponding Values of N and p_1 for Selected Values of d (1 per cent level of significance).

TABLE 1
5% (Roman Type) and 1% (Bold Face Type) Values of N_1 and N_2 for Selected Values of k and d

$k=\frac{N_2}{N_1}$	$d=\lvert p_1-p_2\rvert$																			
	.08		.10		.12		.14		.16		.18		.20		.25		.30			
	N_1	N_2	N_1	N_2	N_1	N_2	N_1	N_2	N_1	N_2	N_1	N_2	N_1	N_2	N_1	N_2	N_1	N_2		
3.0	200	600	128	384	89	267	65	196	50	150	40	119	32	96	20	61	14	43		
	346	**1037**	**221**	**663**	**154**	**461**	**113**	**339**	**86**	**259**	**68**	**205**	**55**	**166**	**35**	**106**	**25**	**74**		
2.8	204	570	130	365	91	253	66	186	51	143	40	113	33	91	21	58	14	41		
	352	**985**	**225**	**630**	**156**	**438**	**115**	**322**	**88**	**246**	**69**	**195**	**56**	**158**	**36**	**101**	**25**	**70**		
2.6	208	540	133	346	92	240	68	176	52	135	41	107	33	86	21	55	15	38		
	359	**933**	**230**	**597**	**159**	**415**	**117**	**305**	**90**	**233**	**71**	**184**	**57**	**149**	**37**	**96**	**26**	**66**		
2.4	213	510	136	327	94	227	69	167	53	128	42	101	34	82	22	52	15	36		
	367	**881**	**235**	**564**	**163**	**392**	**120**	**288**	**92**	**220**	**73**	**174**	**59**	**141**	**38**	**90**	**26**	**63**		
2.2	218	480	140	307	97	213	71	157	55	120	43	95	35	77	22	49	16	34		
	377	**829**	**241**	**531**	**168**	**369**	**123**	**271**	**94**	**207**	**74**	**164**	**60**	**133**	**39**	**85**	**27**	**59**		
2.0	225	450	144	288	100	200	73	147	56	113	44	89	36	72	23	46	16	32		
	389	**778**	**249**	**498**	**173**	**346**	**127**	**254**	**97**	**194**	**77**	**154**	**62**	**124**	**40**	**80**	**28**	**55**		
1.9	229	435	147	279	102	193	75	142	57	109	45	86	37	70	23	45	16	31		
	396	**752**	**253**	**481**	**176**	**334**	**129**	**245**	**99**	**188**	**78**	**148**	**63**	**120**	**41**	**77**	**28**	**53**		
1.8	233	420	149	269	104	187	76	137	58	105	46	83	37	67	24	43	17	30		
	403	**726**	**258**	**464**	**179**	**323**	**132**	**237**	**101**	**181**	**80**	**143**	**65**	**116**	**41**	**74**	**29**	**52**		
1.7	238	405	153	259	106	180	78	132	60	101	47	80	38	65	24	41	17	29		
	412	**700**	**263**	**448**	**183**	**311**	**134**	**228**	**103**	**175**	**81**	**138**	**66**	**112**	**42**	**72**	**29**	**50**		
1.6	244	390	156	250	108	173	80	127	61	98	48	77	39	62	25	40	17	28		
	421	**674**	**270**	**431**	**187**	**299**	**138**	**220**	**105**	**168**	**83**	**133**	**67**	**108**	**43**	**69**	**30**	**48**		
1.5	250	375	160	240	111	167	82	122	63	94	49	74	40	60	26	38	18	27		
	432	**648**	**276**	**415**	**192**	**288**	**141**	**212**	**108**	**162**	**85**	**128**	**69**	**104**	**44**	**66**	**31**	**46**		
1.4	257	360	165	230	114	160	84	118	64	90	51	71	41	58	26	37	18	26		
	444	**622**	**284**	**398**	**197**	**276**	**145**	**203**	**111**	**156**	**88**	**123**	**71**	**100**	**45**	**64**	**32**	**44**		
1.3	265	345	170	221	118	153	87	113	66	86	52	68	42	55	27	35	19	25		
	459	**596**	**293**	**382**	**204**	**265**	**150**	**195**	**115**	**149**	**91**	**118**	**73**	**95**	**47**	**61**	**33**	**42**		
1.2	275	330	176	211	122	147	90	108	69	83	54	65	44	53	28	34	20	23		
	475	**570**	**304**	**365**	**211**	**253**	**155**	**186**	**119**	**143**	**94**	**113**	**76**	**91**	**49**	**58**	**34**	**41**		
1.1	286	315	183	202	127	140	94	103	72	79	57	62	46	50	29	32	20	22		
	495	**544**	**317**	**348**	**220**	**242**	**162**	**178**	**124**	**136**	**98**	**108**	**79**	**87**	**51**	**56**	**35**	**39**		
1.0	300	300	192	192	133	133	98	98	75	75	59	59	48	48	31	31	21	21		
	518	**518**	**332**	**332**	**230**	**230**	**169**	**169**	**130**	**130**	**102**	**102**	**83**	**83**	**53**	**53**	**37**	**37**		

PSYCHOMETRIKA—VOL. 11, NO. 1
MARCH, 1946

A SIMPLE GRAPHICAL METHOD FOR ORTHOGONAL ROTATION OF AXES

WAYNE S. ZIMMERMAN
UNIVERSITY OF SOUTHERN CALIFORNIA

A simple method of making orthogonal rotations by projecting coordinates from one plot to another is presented. An easily constructed apparatus for rapid and accurate work is described in detail. By the technique proposed 12 to 15 rotations of pairs of axes involving 20 test variables can be completed per hour by a practiced technician.

A principal source of resistance against applying the technique of factor analysis is that it is overly time consuming. In particular, rotation to simple structure involves considerable computational work. This is especially true if the method of rotation used involves one rotation at a time with computation of new coordinates after each rotation. The method herein described is one by which the time involved in the plotting and rotating of axes can be reduced to a minimum. By the method proposed, disregarding planning time, approximately 15 pairs of reference axes each involving 20 test variables can be rotated per hour by an experienced operator.

The technique described is a graphic one for orthogonal rotations and makes use of the principle of projection. Coordinates representing the test loadings in one factor are projected horizontally, and coordinates representing the test loadings on the factor with which the first factor is to be rotated are projected vertically, to a common meeting point. Each meeting point represents the position of a test on the two axes to be rotated. Oblique rotations can be introduced after orthogonal transformations have been completed.

The Apparatus

An apparatus for rapid and accurate projection of points can be constructed simply by making use of a 24″ × 30″ drawing board, a 20″ × 30″ sheet of ten-squares-to-an-inch graph paper, a 30″ T square, and an 18″ 30°-60° triangle. Modifications of these instruments to adapt them to this use are described below. See Figure 1.

The graph paper is lined up with the sides of the board by use of

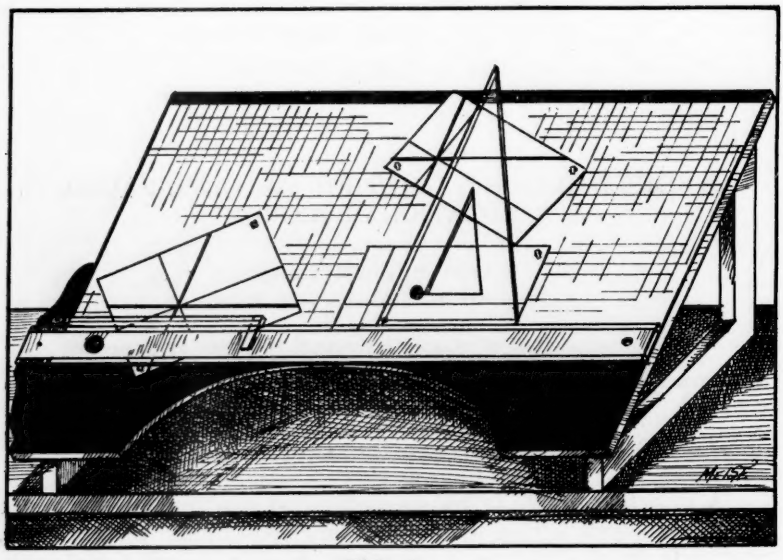


FIG. 1

the T square, and then fastened securely to the surface of the board.* The T square lies horizontally along the board, with the arms snugly against its left-hand edge. Attached to the upper edge of the T square and extending 12 inches from the upper arm is a length of transparent celluloid or plastic on which is scratched a hairline running parallel with the edge of the T square. The hairline can be scratched with a scribe and can be blackened by running a ruling pen filled with India ink down the groove.†

Along the short leg of the triangle, and parallel to the edge, is drawn another hairline, which is a direct continuation of the line on the attachment to the T square. Along the other leg of the triangle, and parallel to the edge, is a third hairline. At the point of intersection of these two lines on the triangle, a hole is drilled large enough to allow a pencil point to be inserted.‡ The hole should be countersunk from the top to facilitate the entrance of the pencil lead

* The graph paper may be mounted on a separate flat surface for initial plotting and final reading purposes. In this case one or two horizontal lines and one or two vertical lines, drawn in appropriate positions on the rotating board, will suffice for guides to locate the plots. By this system the plots can be adjusted more quickly since less perceptual distraction is introduced. For best results the board should be painted white and the lines drawn with black ink.

† All lines should be drawn on the under side to avoid effects of parallax in reading.

‡ Use of automatic pencil is desirable as the hole can be drilled to fit the constant diameter of the lead.

and also slightly countersunk from the bottom to keep lead which may cling to the edge of the hole from smearing the paper when the triangle is moved. Next to the pencil hole an opening can be cut large enough to record the number of the variable represented by the point. A knob attached above and to the right of the hole will be found helpful in making triangle adjustments. A similar knob may be placed on the T square.

Letter-size onionskin paper is used for plotting. To facilitate manipulation of these sheets of paper, it has been found advisable to punch notebook-size holes in the upper-right and lower-left corners and to place over each hole a piece of masking tape. When the sheet is adjusted on the board, pressure over each hole with the thumb or finger will keep the sheet of paper in place. The taped holes will remain adhesive for several days after the removal of the sheet from the board.*

Procedure

To start the rotations the centroid loadings must be plotted from the numerical values. Only nonoverlapping pairs of axes need be plotted to begin with, that is, factors 1 and 2, 3 and 4, and so on. This should be done with great care, using a sharp pencil point. From initial plotting to the completion of the rotational process, the loadings are projected graphically from one pair of sheets to the next sheet. Only at the end are the numerical projections read from the plots to form a table of rotated loadings.

The plotting of centroid axes is simply accomplished by laying a sheet of paper on the board and locating the points by means of the grids on the underlying graph paper, which can be seen plainly. The two axes so plotted may be rotated by turning the sheet until the horizontal and vertical graph lines which intersect at the origin are located in the desired positions relative to the plot on the sheet. The next axes are then drawn using the T square and triangle.

When a rotation is planned with two factors, each of which has already been plotted, e.g., I_5† and III_2 in Figure 2, the procedure is as follows:

(1) In the lower left-hand portion of the board adjust the sheet of paper on which the last rotated plot for one of the factors appeared. It does not matter, of course, what other factor is plotted on the same

* This method of holding the papers in place was suggested by Sgt. J. Boory.
† I_5 represents the fifth rotation of centroid axis I.

sheet. The sheet is so adjusted that the rotated axis of the factor is parallel to the vertical lines on the graph paper beneath. See factor III_2, in Figure 2.

(2) Then, in the upper right-hand portion of the board, adjust the sheet on which the last rotated plot for the second factor appears so that the axis of the factor is parallel to the horizontal lines in the graph paper. See Factor I_5, in Figure 2.

(3) Next adjust a new sheet so that all the test loadings projected horizontally to the right from the lower plot, and projected vertically downward from the upper plot will fall on the paper. See Figure 2.

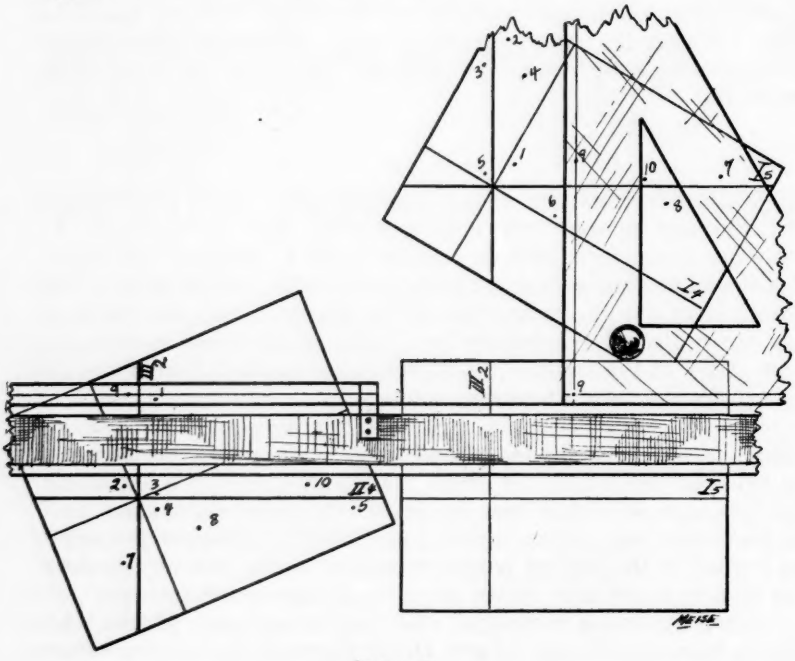


FIG. 2

(4) Using the T square and triangle, draw horizontal and vertical axis lines on the new sheet as continuations of the desired axes on the two previous plots. Draw these lines lightly to facilitate the perceptual discrimination between the plotted axes and the rotated axes which will later be drawn with heavier lines.

(5) Now plot the position of each test on the new sheet. First adjust the hairline on the T square so that it passes through the point

representing the test number which is farthest upward on the lower plot, e.g., variable 9 in Figure 2. Then adjust the vertical hairline of the triangle so that it passes through the point representing the same number on the upper plot.

(6) Insert a pencil in the hole in the triangle and mark the position of the test on the new plot. Record the number of the test.

(7) Move the T square hairline downward until it passes through the next test point on the lower plot, e.g., variable 1 in Figure 2.

(8) Then adjust the vertical hairline of the triangle to pass through the corresponding point on the upper plot.

(9) Mark the position of this test on the new plot and record its number.

(10) Continue in this manner until all test loadings have been projected to the new sheet.

(11) When the rotation has been determined, draw the new axes with a heavier line than those for the original axes.

(12) Label the new axes. On the old sheets indicate by a cross or a check mark the axes which have now been rotated.

With the apparatus currently in use, the error introduced by displacing the pencil point from its geometrically accurate position is very small. A precision-made apparatus could be constructed to operate within almost any limits desired.

Errors in plotting are unlikely, for it is only necessary to be sure that the corresponding numbers are lined up and projected, and that the latest rotations of each factor are used for each new plot. Accuracy depends upon the care with which plots are lined on the board and the hairlines adjusted to the test points.

PSYCHOMETRIKA—VOL. 11, NO. 1
MARCH, 1946

A NEW DESCRIPTIVE STATISTIC: THE PARABOLIC CORRELATION COEFFICIENT

CHARLES C. PETERS
PENNSYLVANIA STATE COLLEGE

This article proposes a new descriptive statistic related to the second order parabola in the same manner in which the familiar correlation coefficient is related to the regression coefficient. The parabolic r describes in standard terms simultaneously the general trend of the regression and the extent and nature of its curvilinearity, and is relatively easy to compute and easy to communicate. Formulas for its standard error and its limits are derived and it is applied to a number of regressions.

It is a well-known fact that the correlation coefficient expresses the slope of a *straight* regression line on a squeezed scale, and that it is an inappropriate and misleading descriptive statistic where the relation between two sets of paired variates is not rectilinear. Nevertheless it has been recklessly used without regard to the rectilinearity of the regression. That has been in part due to the absence of any convenient descriptive statistic for curvilinear correlational relationships. We have long had, of course, the equations of parabolas of various orders and also a number of other forms of curves. But these have not been in the form of standard units so as to make their descriptions of the regressions readily communicable. It is the purpose of this paper to describe a new statistic related to the equation of the second-order parabola in the same manner as the correlation coefficient is related to the regression coefficient, to give machinery for translating it into standard meaning, and to develop for it a standard error formula and a proof of its limits. I have named this statistic *the parabolic correlation coefficient.* It provides a dual description of the regression: (1) an index of the slope (i.s.) exactly equal in numerical value to the correlation coefficient; and (2) an index of corvilinearity (i.c.) The index of slope is the first derivative of the equation of the second-order parabola evaluated at a certain point to be described in this article, while the index of curvilinearity is one half of the second derivative, each adjusted to a standard meaning by a suitable squeezing of the scale.

The equation for the second degree parabola is

$$Y = a + bX + cX^2. \tag{1}$$

We wish to get values for the a, b, and c such that the sum of the squares of the errors made in predicting Y-scores with this equation will be a minimum. In the case of any individual with an X-score in the class i the error would be

$$(Y_i - \hat{Y}_i) = Y_i - (a + bX_i + cX_i^2).$$

We want the sum of the squares of these errors to be a minimum. Hence we square the expression at the right and sum it termwise. But we shall write the terms in a different order, merely for convenience and, for the sake of simplicity, shall drop the subscripts. The sum of the squares of the errors is:

$$\begin{aligned}\Sigma Y^2 + c^2\Sigma X^4 + b^2\Sigma X^2 + Na^2 - 2c\Sigma X^2Y - 2b\Sigma XY \\ - 2a\Sigma Y + 2ac\Sigma X^2 + 2ab\Sigma X + 2bc\Sigma X^3.\end{aligned}$$

In order to determine values for a, b, and c that will make this expression for the sum of the squares of the errors a minimum, we must equate to zero the partial derivatives with respect to these three terms. These partial derivatives are in order (each equated to zero and each divided through by 2):

$$\begin{aligned} c\Sigma X^4 + b\Sigma X^3 + a\Sigma X^2 - \Sigma X^2Y &= 0, \\ c\Sigma X^3 + b\Sigma X^2 + a\Sigma X - \Sigma XY &= 0, \\ c\Sigma X^2 + b\Sigma X + Na - \Sigma Y &= 0. \end{aligned}$$

This set of simultaneous equations must be solved for the three unknowns, a, b, and c. A number of different outcomes, all completely equivalent in value, are possible according to the method of handling the algebra. One of the most convenient solutions is the following:

$$c = \frac{(\Sigma X^2 \Sigma Y - N\Sigma X^2 Y)(N\Sigma X^2 - [\Sigma X]^2) + (N\Sigma XY - \Sigma X \cdot \Sigma Y)(N\Sigma X^3 - \Sigma X^2 \Sigma X)}{(N\Sigma X^3 - \Sigma X^2 \Sigma X)^2 + ([\Sigma X^2]^2 - N\Sigma X^4)(N\Sigma X^2 - [\Sigma X]^2)}. \quad (2)$$

$$b = \frac{(N\Sigma XY - \Sigma X \cdot \Sigma Y) - c(N\Sigma X^3 - \Sigma X^2 \Sigma X)}{N\Sigma X^2 - [\Sigma X]^2}. \quad (3)$$

$$a = \frac{\Sigma Y - c\Sigma X^2 - b\Sigma X}{N}. \quad (4)$$

These are the coefficients which are now to be put in equation (1) for predicting Y scores.

But so far we have an operative tool, not a descriptive statistic. We shall turn next to getting the latter. The reader should get in mind a conventional correlation chart laying out the independent scores on the X-axis and the dependent scores on the Y-axis. We are concerned with two features about such outlay: (1) What is the general trend of the Y scores as we go up along the X-axis? and (2) To what extent is the regression line curved rather than straight, and in what direction is it curved? We can get an account of the first feature from the first derivative of the equation of the parabola and an account of the second feature from the second derivative.

The first derivative of equation (1) is $b + 2cX$. This gives the slope of a tangent to the curve at any point at which the derivative is evaluated. Its value is dependent upon its place along the X-axis, as is evidenced by the fact that it contains an X. If we evaluate it near the middle of the X-distribution, we get a tangent that will show the general trend pretty well for the whole distribution; and if, in addition, we make a proper adjustment in our scales, we can have a description of the general trend that compares in value with the coefficient of correlation. We shall adjust our scales by stretching or squeezing them so as to make the units equal in terms of variability on the two axes. This we accomplish by multiplying any increments along the X-axis (here the first derivative) by σ_x/σ_y. But we shall put this in summation form in the first term since we have all the other terms in summation form, but shall, for convenience in writing, let it stand in σ form in the other two terms. So re-writing our first derivative thus multiplied by the standard deviation ratio and separating it into three terms, and at the same time putting the $2c$ over the common denominator, and also substituting NM_x for $\sum X$, we have

$$i.s. = \frac{N\sum XY - \sum X \cdot \sum Y}{\sqrt{(N\sum X^2 - [\sum X]^2)\,(N\sum Y^2 - [\sum Y]^2)}} - \frac{c(N\sum X^2 - NM_x\sum X^2)}{N^2\,\sigma_x\,\sigma_y} + \frac{2cXN^2\,\sigma_x^2}{N^2\,\sigma_x\,\sigma_y}. \tag{5}$$

Inspection of the above expression will show that, if we evaluate it where the second and third terms aggregate exactly zero, we shall have left precisely the formula for the Pearsonian correlation coefficient. Since the denominators are the same, that will be where

$$2cXN^2\sigma_x^2 = c(N\Sigma X^3 - NM_x\Sigma X^2).$$

Solving this for X, we get

$$X = \frac{\Sigma X^3 - M_x\Sigma X^2}{2N\sigma_x^2}. \tag{6}$$

So if we evaluate the first derivative at the position on the X-axis where X has the value indicated in (6) on our adjusted scale, we shall always get for this part of our descriptive statistic exactly the numerical value of the coefficient of correlation. Investigation of this expression will show that it reduces to the mean in the case of a symmetrical distribution of the X-variable; but we can easily get for the needed critical point at which to evaluate it a more generalized value.

The expression as it stands in (6) is in score form. In order to investigate its properties we shall transform it to deviation form. The capital X represents a score and the lower case x a deviation from the mean of the X's. So

$$X = x + M_x.$$

Hence

$$\begin{aligned} \Sigma X^3 &= \Sigma x^3 + 3M_x\Sigma x^2 + 3M_x^2\Sigma x + NM_x^3; \\ M_x\Sigma X^2 &= M_x\Sigma x^2 + 2M_x^2\Sigma x + NM_x^3. \end{aligned} \tag{7}$$

Subtracting the second equation in (7) from the first to get the numerator of (5), remembering that no matter what the shape of the distribution $\Sigma x = 0$, and simplifying, we have

$$X = \frac{2M_x\Sigma x^2 + \Sigma x^3}{2N\sigma_x^2} = M_x + \tfrac{1}{2}\sigma_x\sqrt{\frac{(\Sigma x^3)^2}{N^2\sigma^6}}. \tag{8}$$

The term under the radical above is, in the Pearson terminology, β_1. But because the sum of the odd powers may have either the plus or minus sign before squaring, and because we wish to preserve the effect of the sign, we shall write the equation as follows:

$$X = M_x \pm \tfrac{1}{2}\sigma_x\sqrt{\beta_1}. \tag{9}$$

For a normal distribution, or for any symmetrical distribution, β_1 is zero. So, in order to get for this part of our descriptive statistic a value exactly equal to the coefficient of correlation, we must evaluate it exactly at the mean if the X-distribution is normal or otherwise symmetrical. If the distribution is not symmetrical, the critical point would be near the mean but a little off to the extent involved in (9). But in the computational work the worker need not be concerned about that. If he will merely compute

$$i.s. = \frac{N\Sigma XY - \Sigma X \cdot \Sigma Y}{\sqrt{(N\Sigma X^2 - [\Sigma X]^2)\,(N\Sigma Y^2 - [\Sigma Y]^2)}}$$

it will follow as an automatic consequence that he has evaluated the derivative at the position indicated in (9) and that he has the slope of a tangent to the curve exactly parallel to the fitted line the slope of which would be the coefficient of correlation. It would, of course, be plausible always to evaluate the derivative at the mean, but then our value for the *i. s.* would not always have a familiar meaning. It would seem worth while to sacrifice that point for the practical advantage of continuing to carry in our parabolic correlation coefficient an element exactly equal to the well-known Pearsonian r.

The above discussion provided an index of slope for our descriptive statistic. We need for our second element an index of the curvilinearity. This requires the second derivative. The second derivative of equation (1) is $2c$. But for certain practical reasons discussed below we shall take for our *i.c.* half of that derivative, namely c. But we also want it expressed in terms of our adjusted scales in order that it may have a standard meaning for all types of distributions to which it is applied. That means that we must multiply it by the ratio σ_x^2/σ_y. When this multiplier is put in summation terms and used with equation (2) we get for the *i.c.* equation (10a) below. When used with another of the three possible solutions for c, we get equation (10b) below. In practice it is desirable to employ both of these formulas as an arithmetic check. They should give identical outcomes.

$$i.c. = \frac{(\Sigma X^2 \Sigma Y - N\Sigma X^2 Y)(N\Sigma X^2 - [\Sigma X]^2) + (N\Sigma XY - \Sigma X \cdot \Sigma Y)(N\Sigma X^3 - \Sigma X^2 \Sigma X)}{(N\Sigma X^3 - \Sigma X^2 \Sigma X)^2 + ([\Sigma X^2]^2 - N\Sigma X^4)(N\Sigma X^2 - [\Sigma X]^2)} \cdot \frac{(N\Sigma X^2 - [\Sigma X]^2)}{N\sqrt{N\Sigma Y^2 - [\Sigma Y]^2}}. \quad (10a)$$

$$i.c. = \frac{(\Sigma X^2 Y \cdot \Sigma X - \Sigma X^2 \Sigma XY)(N\Sigma X^2 - [\Sigma X]^2) + (N\Sigma XY - \Sigma X \cdot \Sigma Y)([\Sigma X^2]^2 - \Sigma X^3 \Sigma X)}{(\Sigma X^4 \Sigma X - \Sigma X^3 \Sigma X^2)(N\Sigma X^2 - [\Sigma X]^2) + (N\Sigma X^3 - \Sigma X^2 \Sigma X)([\Sigma X^2]^2 - \Sigma X^3 \Sigma X)} \cdot \frac{(N\Sigma X^2 - [\Sigma X]^2)}{N\sqrt{N\Sigma Y^2 - [\Sigma Y]^2}}. \quad (10b)$$

The reader may be alarmed at the elaborateness of this formula for the index of curvilinearity. Its application does, it must be admitted, require some work. But it will be observed that the same quantities recur in it, in consequence of which the labor is less than would at first appear. The quantities in all of the parentheses lend themselves to very economical procedures on a calculating machine. There are needed, besides the N, the following moments and product moments:

$$\Sigma X$$
$$\Sigma Y$$
$$\Sigma X^2$$
$$\Sigma Y^2$$
$$\Sigma XY$$
$$\Sigma X^2 Y$$
$$\Sigma X^3$$
$$\Sigma X^4$$

The first five of these are needed in the computation of a coefficient of correlation; only the last three are additional. They can be very easily and quickly obtained when working with a correlation chart, though no correlation chart is really needed for this statistic as it is for the correlation ratio and the Blakeman Test. Getting them on a calculating machine requires going through the scores one additional time as compared with the correlation coefficient, and the total time to calculate a parabolic r from raw scores on a calculating machine is about twice that required for a Pearsonian coefficient of correlation!* The needed moments and product moments can also be obtained on the I.B.M. machines, but this requires that the X^2 values be punched in the cards as well as the X values.

So our parabolic correlation coefficient will contain always two terms, one giving in familiar form an index of the general slope and the other giving an index of the form and the degree of curvilinearity. If the regression is really rectilinear, the second element will be zero and the parabolic correlation coefficient will automatically reduce to the Pearsonian r. In any case the parabolic correlation coefficient will automatically carry notice of the degree of departure from rectilinearity, and we can, if we wish, test that departure for statistical significance by the method explained below. A parabolic r of, say, $-.163 - .167$ means that there is in our relation a moderate over-all downward trend (i.s. $= -.163$) and there is an upward hump of moderate amount in the curve (i.c. $= -.167$). A parabolic r of $-.163 +$

* Convenient worksheets for the parabolic correlation coefficient may be obtained from the Department of Psychology and Education of the Pennsylvania State College at one cent each.

.675 would mean that there is a general downward trend and also a distinct downward hump. Thus the nature of the regression can be visualized from the statement of the parabolic r without the need for a correlation chart.

What are the maximum and minimum limits between which our index of curvilinearity lies? My proof of these limits rests upon two postulates: the *i.c.* can be a maximum or a minimum only (1) when the index of slope is zero, and (2) when the regression law operates perfectly to place each Y-value exactly on the curve when allocated to its appropriate position along the X-axis. Hence, where σ_y is the standard deviation of the actual Y-values and σ_{y_c} is the standard deviation of the Y's on the regression line,

$$\sigma_y = \sigma_{y_c} = \left[\frac{\Sigma(a + bX + cX^2)^2}{N} - M_y{}^2\right]^{\frac{1}{2}}.$$

When this expression is evaluated it results in moments in terms of ΣX^4, ΣX^3, ΣX^2, and ΣX. These are in score form; they can be more easily simplified if transformed to deviation form. Let $X = x + M_x$, where the lower case x is a deviation from the mean of the X's. Raising this to the 4th power, summing, and dividing by the N of the denominator above yields terms in $\Sigma x^4/N$, $\Sigma x^3/N$, $\Sigma x^2/N$, and $\Sigma x/N$. The last two are, of course, $\sigma_x{}^2$ and zero. The first is $\beta_2\sigma_x{}^4$ and the second is $\sigma_x{}^2\sqrt{\beta_1}$. The X^3 can be correspondingly treated. When these values are substituted, and also values for a and for b from equations (3) and (4), and cognizance is taken of the fact that the *i.s.* must be zero, the whole expression reduces to

$$\sigma_y = c\sigma^2\sqrt{\beta_2 - \beta_1 - 1}.$$

The algebra involved in this is straightforward substitution but to carry it through requires several pages, which I do not deem it necessary to consume here, because anyone who is interested can carry through the proof himself from the above hints. We now substitute this value for the denominator in formula (10). The $\sigma_x{}^2$ is canceled by the numerator of the same fraction. The c standing in the substituted denominator here is canceled by the lengthy fraction constituting the left-hand factor of equation (10), which is exactly c. Hence the whole reduces to the very simple form

$$i.c.^{\text{max.}} = \frac{1}{\sqrt{\beta_2 - \beta_1 - 1}}.$$

The minimum value is had by using the negative sign with the square root. The β's here are the Pearsonian measures of kurtosis and of skewness, respectively, of the X-variable (the independent variable).

For a normal distribution ($\beta_2 = 3$ and $\beta_1 = 0$) the maximum limit of the *i.c.* would be .707. . . . But for a rather platykurtic distribution the *i.c.* could reach ± 1.00 or even, though rarely in practice, exceed those limits. It is because of the desirability of pairing two elements in our parabolic r that are confined to substantially the same limits that we chose half of the second derivative for our *i.c.* instead of the second derivative as it stood. It is, perhaps, a further justification that the *i.c.* then becomes just the c of the parabolic equation, though on a squeezed scale.

Statistical workers accustomed to the conventional testing for reliability will ask for standard error formulas for this statistic. Since the parabolic correlation coefficient is merely a descriptive statistic and has no functional meaning as a whole, we do not need a standard error formula for it as a whole, though one could be developed if wanted. We are interested only in testing the reliability of its two parts. For testing the hypothesis of a true slope of zero, the standard error of the index of slope where we have evaluated it is merely that long familiar for testing the corresponding hypothesis in the correlation coefficient:

$$S.E._{i.s.} = \frac{1}{\sqrt{N-1}}.$$

But we desire a formula for the standard error of the index of curvilinearity. Deming* has shown that the sampling variance of the coefficient of X^2 in a second-order parabola, c, is

$$S.E._{c}{}^2 = \frac{\sigma_y{}^2}{(N-3)\,(\sum x^4/N - \sigma^4)}, \tag{11}$$

where $\sigma_y{}^2$ is the variance of the Y-array in the sample and the x's in the denominator are deviations from the X-mean. But that form would be too inconvenient in use. We want a form that can be easily applied mentally as soon as we hear the index of curvilinearity and know the N. We can simplify (11) by assuming normality of distribution in the X-variable. Normality is not a very violent assumption in the X-variable, although it would be for the Y-variable if there is marked curvilinearity.

$\sum x^4/N\sigma_x{}^4$ is β_2 in the Pearson terminology. So $\sum x^4/N = \beta_2\,\sigma_x{}^4$. But in a normal distribution, $\beta_2 = 3$. So the first term in the parentheses in (11) equals $3\sigma^4$. This combines with the second member to make $2\sigma^4$. Making this substitution and taking the square root, we have for the standard error of the coefficient c,

* Deming, W. E. Statistical adjustment of data. John Wiley and Sons, 1943.

$$S.E._{.c} = \frac{\sigma_y}{\sigma_x^2 \sqrt{2(N-3)}}. \tag{12}$$

But our index of curvilinearity is $\frac{\sigma_x^2}{\sigma_y} c$. On the principle that $\sigma_{ax} = a\sigma_x$, we would then have for the standard error of our index of curvilinearity:

$$S.E._{i.c.} = \frac{\sigma_x^2}{\sigma_y} \cdot \frac{\sigma_y}{\sigma_x^2 \sqrt{2(N-3)}} = \frac{1}{\sqrt{2(N-3)}}. \tag{13}$$

In mathematical theory this has the same meaning and use as the standard error of any other statistic, and is to be divided into the c to get the standard error ratio. But in this sort of situation, as in many others, the worker should be warned against the customary acceptance of the hypothesis that there may be no true curvilinearity until the standard error ratio reaches the five per cent or the one per cent level. A test of statistical significance is properly employed to caution the worker against proceeding to action on the basis of his findings when there still remains considerable possibility that the difference tested may have arisen by chance fluctuation in sampling. In an experiment on methods of teaching, for example, one should withhold confident action even when the odds are 90 to 10 that the true difference is on the side on which he found it, for they are still 10 to 90 that the true difference might turn out to be on the other side. But in respect to taking warning that there may be true curvilinearity in our regression the danger of over-hasty action lies on the opposite side; here the unwarrantedly hasty action would consist in disregarding the evidence with low probability that there may be true curvilinearity and proceeding to treat the regression as rectilinear. So, whereas we hold forth for a standard error ratio of two or two and a half before we claim statistical significance in an experiment, we would better take warning from a standard error ratio of one in this connection (as, indeed, in a number of other situations of like character). The actual numerical size of the *i.c.* is more important that its standard error ratio. An *i.c.* of .06 or .10 or higher arithmetically shows an appreciable amount of curvilinearity.

In the table below are reported a number of parabolic *r's*. These are not a random sample nor are they selected as only the ones showing curvilinearity; they are ones computed from data in which it seemed plausible that we might find curvilinearity. In order to appreciate the phenomena involved in them, it may be well to approach the table by speculating about what might be expected to be the relation. A number of the parabolic *r*'s are ones computed between scores in subject fields made at a single long examination by

Parabolic Correlation Coefficients Between Indicated Pairs of Variates

Variates	*i.s.*	*i.c.*	$S.E._{i.c.}$	*Ratio*
History on mathematics	+ .08	—.168	.067	2.49
Percentage of votes for Willkie by counties in Pa. on density of population	—.291	+.208	.089	2.34
Percentage of votes for Willkie in 1940 by Counties in Ohio on density of population	—.402	+.011	.077	.14
Otis Classification test on age at eighth-grade level	—.163	—.167	.045	3.71
Literature on grade-point average	+.347	—.138	.067	2.04
Literature on mathematics	+.113	—.125	.067	1.85
Belief in superstitions on I.Q.	—.199	—.054	.021	2.57
Belief in superstitions on no. of magazines taken in home	—.145	—.037	.021	1.76
Belief in superstitions on no. of semesters taken in science	—.101	+.033	.021	1.50
Impulsiveness on success in nursing	—.265	—.125	.072	1.74
Indecision on success in nursing	—.140	—.054	.072	.75
Worry on success in nursing	—.027	—.041	.072	.57
Literature on science	+.408	—.111	.067	1.64
Grade-point average on math.	+.321	—.111	.067	1.64
Mathematics on literature	+.113	—.109	.067	1.61
English on grade-point average	+.528	—.081	.067	1.17
Washburne control scores on academic achievement ratio	—.034	+.093	.041	2.27
Washburne alienation scores on academic achievement	+.126	+.052	.041	1.27
Happiness scores on academic achievement	+.042	+.053	.041	1.29
Number of siblings on academic achievement ratio	—.091	—.006	.041	.15
Science on general intelligence	+.306	+.068	.067	1.01
English on mathematics	+.162	—.062	.067	.92
English on history	+.463	—.062	.067	.90
Mathematics on history	+.082	+.022	.067	.32
Moore-Nell academic achievement test on behavior inventory	+.033	—.010	.051	.19
English total on intelligence	+.728	+.022	.067	.33

113 sophomores at the Pennsylvania State College on the Carnegie Foundation Achievement Tests. As an individual increases his ability in mathematics, what is likely to happen in his measured abilities in history? The two abilities may be expected to increase together at the low end on account of the common elements of reading ability, general intelligence, etc. But in the later stages those persons who go in much for mathematics are likely not to be going in as much for history, and hence at the upper levels the curve relating the two could be expected to bend downward from the straight regression line that

would have fitted the early stages. Thus there would be a humped up curve for the regression line, represented by a negative index of curvilinearity. That is exactly what inspection of the first line in the table shows to be the fact. Nearly all the other parabolic r's bear out similarly the hypotheses one would set up for them as plausible. In the table the first term in each pair denotes the dependent variable and the second term the independent variable. Thus *history on mathematics* shows the manner in which the history scores regress as one goes up the scale of mathematics scores.

We have talked in terms of a second-order parabola. Parabolas of higher order could, of course, also be used and the descriptive statistic would need to have as many terms as the order of the parabola. But beyond the second order the computational work would greatly increase and the difficulty of communicating the results and of visualizing their meaning would likewise increase. It would also be possible to set up descriptive statistics in terms of other curves, such as the Gompertz or other exponential curves. But they also would be more awkward to compute and to communicate. Moreover, the second-order parabola will serve pretty well to detect and describe curvilinearity through the ranges through which we usually measure. We are likely to be measuring growth some time after its initial period, when it is in the upper bend of the S of the Gompertz curve; and in relations which are fitted by curves of the form $Y = e^{ax}$, we are also likely to be working within a range in which the second-order parabola will not fit too badly. But for actual prediction rather than rough description we should resort to the fitting of the needed type of curve. It would thus appear that the constants of a second-order parabola will make at least a satisfactory start in a more adequate description of regression than the sole resort to correlation coefficients upon which we have hitherto relied. Although our experience suggests that a disturbing amount of curvilinearity in regression is far from the rule, it does occur sufficiently frequently that the careful researcher will wish to be on the lookout for it. It is particularly likely to be missed when correlation coefficients are run on a calculating machine, or on the I.B.M. machines from ungrouped scores, where, consequently, there is no inspectional warning of the curvilinearity.